(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number
WO 01/18057 A2

(51) International Patent Classification[7]: C07K 14/81, C12N 15/15, 15/63, C07K 16/18, G01N 33/50

(21) International Application Number: PCT/GB00/03448

(22) International Filing Date: 7 September 2000 (07.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9921133.6 7 September 1999 (07.09.1999) GB

(71) Applicant (*for all designated States except US*): CAMBRIDGE UNIVERSITY TECHNICAL SERVICES LIMITED [GB/GB]; University of Cambridge, 20 Trumpington Street, Cambridge CB2 1QA (GB).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): GUBB, David, Christopher [GB/GB]; University of Cambridge, Department of Genetics, Downing Street, Cambridge CB2 3EH (GB). REICHHART, Jean-Marc [FR/FR]; § UPR 9022 C.N.R.S., Institut de Biologie Moléculaire et Cellulaire, 15, rue René Descartes, F-67084 Strasbourg (FR).

(74) Agents: KIDDLE, Simon, J. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MATERIALS AND METHODS RELATING TO THE IDENTIFICATION AND MUTATION OF A *DROSOPHILA* GENE, AND USES THEREOF

(57) Abstract: Tree related *Drosophila* genes have been characterised, and their nucleic acid and encoded amino acid sequences are provided herein. Developmental expression profiles and *in situ* hybridisation patterns for the encoded proteins have been obtained. Transgenic experiments on mutant forms of the genes have demonstrated a link between mutations in one of the genes and the occurrence of a characteristic mutant phenotype of Drosophila flies. Materials, methods and uses based on this discovery are also described.

WO 01/18057 A2

Materials and Methods Relating to the Identification and Mutation of a *Drosophila* Gene, and Uses Thereof.

Field of the Invention

Broadly, the present invention relates to the characterisation of three related *Drosophila* genes, and to materials and methods deriving from these findings. In particular, the present invention relates to nucleic acid molecules comprising these genes, their encoded polypeptides, and the use of such nucleic acids, polypeptides and mutants thereof.

Background of the Invention

The polypeptide product of the *Drosophila Spn43Ac* gene functions to inhibit serine protease activity at the extracellular head of a cytoplasmic signalling pathway. This cytoplasmic pathway, which is triggered by activation of the transmembrane receptor Toll, controls the production of, *inter alia*, anti-fungal and anti-viral polypeptides, and is shown in Figure 1.

The *Toll* gene was initially identified as a maternal effect gene involved in control of the dorso-ventral pattern of the *Drosophila melanogaster* embryo. *Toll* was subsequently shown to encode a transmembrane receptor, the activation of which triggers a cytoplasmic signalling pathway controlling the activity of a transcription factor called Dorsal. Dorsal, like NF-κB, is a member of the Rel family of inducible transactivators, a group of homo- or heterodimeric proteins which play a major role in the regulation of many mammalian immune-response genes.

Significantly, the cytoplasmic domain of Toll shares marked sequence similarities with the cytoplasmic domain of the interleukin-1 receptor (IL-1R). However, both receptors differ by their extracellular domains; that of Toll

consists predominantly of leucine-rich repeats, whereas the corresponding region of the IL-1R has immunoglobulin-like domains.

It is known that Toll and all members of the downstream signalling cascade are expressed in the fat body (immune-responsive tissue) of *Drosophila*. The expression of Toll and other members of the cascade is upregulated following immune challenge. Studies in flies have also established that the Toll signalling cascade controls anti-fungal host defence and, in particular, the expression of the anti-fungal peptide drosomycin in the fat body cells. These results prompted the search for, and discovery of, mammalian Toll homologues expressed in cells of the immune system. The human Toll homologue was shown to activate signal transduction via NF-κB, leading to the production of pro-inflammatory cytokines. Four additional Toll-like receptors were subsequently cloned in mammals.

In parallel to the analysis of the downstream events initiated by activation of the Toll transmembrane receptors, studies were undertaken to identify ligands for the Toll receptors. A first report by Yang et al., carried out on human epithelial cells engineered to express the human Toll-like receptor 2 (TLR 2), indicated that Toll confers sensitivity and selectivity to lipopolysaccharide (LPS) in the presence of LPS-Binding Protein (LBP) and CD14. In addition, it was shown that the response involves activation of NF-κB.

It has been proposed that Toll-like receptors detect molecular structures which are conserved in microbes. This proposal is strengthened by recent studies which used positional cloning to identify the murine gene involved in LPS-sensitivity as a homologue of the human Toll-like receptor 4.

In *Drosophila*, genetic data implicates the *spaetzle* gene product as being a ligand for Toll in embryonic dorso-ventral patterning. Spaetzle is a 40 kDa protein, which shares the cysteine-knot arrangement of several mammalian cytokines and nerve growth factor, NGF. It is cleaved to an active low molecular weight form, corresponding to the 106 C-terminal residues, by a cascade of serine proteases which have been defined, again by genetic analysis, as the products of the genes *gastrulation defective, snake* and *easter*. In previous studies, the present inventors have noted that the induction of drosomycin is dramatically affected in flies mutant for *spaetzle*. The same effect is not, however, observed in flies mutant for genes encoding the upstream serine proteases of the embryonic cascade. These data indicate that, during the immune response, either the Spaetzle protein can activate the Toll receptor without undergoing a proteolytic cleavage, or that other protease genes can substitute for Gastrulation defective, Snake and Easter.

A number of ethyl methane sulphate (EMS)-induced recessive mutations which lead to inappropriate melanisation in *Drosophila* have been identified. Amorphic mutants of the *Drosophila Spn43Ac* gene, which has been named *necrotic* (*nec*), die in the late pupal stage as pharate adults, or hatch as weak, but relatively normal-looking, flies. The adult flies develop black melanized spots on the body and leg joints and die within 24 hours of eclosion. Figure 3 shows a schematic map of the *Spn43A* region on the right arm of *Drosophila* chromosome 2. A number of transcripts mapping to this region have now been identified in the process of defining the tissue polarity gene *prickle* (*pk*). The *Spn43Aa, -b* and *-c* loci have been identified as a cluster of three small transcripts, two within the 5'

intron, and one just proximal to *pk*, that are themselves not part of the *pk* gene. These transcripts encode a set of putative serine proteinase inhibitors, or serpins (SERine Proteinase Inhibitors).

Serpins form a divergent group of proteins that have been found in plants, birds, mammals and viruses. They bind as competitive substrates to the active sites of their target proteases to block the protease activity; in binding, serpins may themselves be cleaved. In the absence of serpins, serine proteases may cleave their normal substrate to produce the active form of the substrate.

In mammals, a variety of proteolytic cascades, including blood coagulation, the complement reaction and the inflammatory response are regulated in this way. Invertebrate serpins are less well characterised. Several serpins have been isolated in *Manduca sexta* and two in *Drosophila melanogaster*, but no genetic functions have been identified with these transcripts.

The present inventors have now achieved the cloning and sequencing of the *Drosophila Spn43Aa*, *Spn43Ab*, *Spn43Ac* transcripts. Nucleotide and deduced amino acid sequences are disclosed herein. Figure 2 shows the nucleotide sequence of the Spn43Ac gene. Figure 9 shows the predicted amino acid sequences of the polypeptides encoded by the *Spn43Aa*, *Spn43Ab* and *Spn43Ac* genes. The inventors have also obtained developmental expression profiles and imaginal disc *in situ* hybridisation patterns for the three serpins.

The present inventors now show that in transgenic experiments, *Spn43Ac* (*necrotic*) rescues the *nec* melanotic phenotype, indicating that mutations in the *Spn43Ac* gene are responsible for the appearance of necrotic brown spots throughout the body of the fly. They also show that

mutation in the serpin *Spn43Ac* gene leads to constitutive expression of the anti-fungal peptide gene *drosomycin*, but not of anti-bacterial peptide genes. Significantly, this phenotype is dependent on the function of the *spaetzle* and *Toll* genes, since it is abolished in a *spaetzle*-deficient and *Toll*-deficient background. The inventors provide evidence that an immune challenge leads to the rapid cleavage of the Spaetzle protein to its low molecular weight active form, and show that, in *necrotic* mutants, both the uncleaved and cleaved forms are constitutively present. Furthermore, they show that the expression of the *Spn43Ac* gene is itself regulated during an immune response *via* the Toll signalling cascade.

These results have led the present inventors to conclude that, in the *Drosophila* host defence, Toll does not directly function as a pathogen pattern recognition receptor, but instead responds to the endogenous product of a proteolytic cascade in the blood of the fly. In view of the similarities between members of the coagulation cascade in horseshoe crabs, the prophenoloxidase-activating cascade in crustaceans and the proteolytic cascade which is regulated by the *Spn43Ac serpin* gene, the inventors teach that the latter is also activated by upstream proteins capable both of binding microbial motifs (pattern recognition) and of initiating a proteolytic cascade upon binding these motifs.

Summary of the Invention

Thus, in one aspect, the present invention provides an isolated polypeptide which comprises an amino acid sequence as set out in Figure 9.

In a further aspect, the present invention provides an isolated polypeptide which is a mutant, variant, derivative

or allele of the above polypeptide. A polypeptide which is a mutant, variant, derivative or allele may have an amino acid sequence which differs from that given in Figure 9 by one or more of an addition, substitution, deletion and insertion of one or more amino acids. The mutant, variant, derivative or allele will retain a biological property of a serpin polypeptide as shown in Figure 9. This property may, for instance, be common immunoreactivity such as cross reactivity with an antibody, or an ability to inhibit serine protease activity. However, as is described in detail below, the present invention also concerns mutant forms of serpin polypeptides which have an impaired ability to inhibit serine protease activity. A polypeptide which is an amino acid sequence mutant, variant, derivative or allele of any one of the above polypeptides may comprise an amino acid sequence which shares greater than about 20% sequence identity with a sequence shown in Figure 9, greater than about 30%, greater than about 40%, greater than about 50%, greater than about 60%, greater than about 70%, greater than about 80%, greater than about 90% or greater than about 95%. Particular amino acid sequence variants may differ from those shown in Figure 9 by insertion, addition, substitution or deletion of 1 amino acid, 2, 3, 4, 5-10, 10-20 20-30, 30-50, 50-100, 100-150, or more than 150 amino acids.

The present invention also includes active portions, fragments, chemical derivatives and functional mimetics of the serpin polypeptides of the invention. An "active portion" of a serpin polypeptide means a peptide which is less than full length serpin polypeptide, but which retains a biological activity, such as ability to inhibit serine protease activity. A "fragment" of the serpin polypeptide means a stretch of amino acid residues of at least about

five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids. Fragments according to the invention may comprise antigenic determinants or epitopes useful for raising antibodies to the full-length serpins. A "chemical derivative" of the serpin polypeptide or a fragment thereof means a polypeptide in which one or more of the amino acid residues are chemically altered eg acetylated. "Functional mimetic" means a substance which may not contain an active portion of the serpin amino acid sequence, and probably is not a peptide at all, but which retains the essential biological activity of a natural serpin polypeptide (e.g. ability to inhibit serine protease activity). The design and screening of candidate mimetics is described below.

In one embodiment, the active portion, fragment, derivative or functional mimetic comprises a serine protease-binding site e.g. a serine protease-binding domain of the serpin. Since such a domain represents the binding domain responsible for the interaction between serpins and serine proteases, it can be used in methods of screening for e.g. agents which act as modulators e.g. inhibitors of the serpin-serine protease interaction (see below).

The present invention further provides in another embodiment, an isolated nucleic acid which has a nucleotide sequence which encodes a polypeptide or peptide as described above. Also provided is an isolated nucleic acid which has a nucleotide sequence complementary to an encoding nucleotide sequence as stated above. In one embodiment, the nucleic acid comprises DNA having a nucleotide sequence as shown in Figure 2, or a portion thereof.

Further provided is an antisense oligonucleotide having a sequence complementary to a nucleic acid as provided herein. The present invention also provides the use of a nucleic acid as defined above in the design of antisense oligonucleotides to restrict serpin expression in a population of cells, e.g. phosphorothiolated or chloresterol-linked oligonucleotides which can facilitate internalization and stabilization of the oligonucleotides. The person skilled in the art can readily screen fragments of various sizes and from various parts of the serpin-encoding sequence to optimise the level of anti-sense inhibition. For hybridization the antisense oligonucleotide needs to have sufficient complementarity or similarity to target sequence; complete sequence identity is not essential. The antisense molecule may therefore differ in one or more nucleotides from the target serpin nucleotide sequence.

Also provided, in further aspects, is a vector or construct comprising any of the nucleic acid or oligonucleotide molecules described above. Further provided is a host cell comprising such a vector or construct. The host cell may be any cell typically used in an expression system eg E.coli. The host cell may be a *Drosophila* cell. The host cell may be used to produce high levels of a serpin polypeptide.

In a further aspect, there is provided a method for producing a serpin polypeptide, comprising culturing a host cell described above under conditions suitable for expression of the serpin polypeptide. In a further step, the polypeptide may be recovered from the cell culture.

A serpin polypeptide or peptide according to the present invention may be used as an immunogen or otherwise in obtaining specific antibodies. Antibodies are useful in

purification and other manipulation of polypeptides and peptides, diagnostic screening and therapeutic contexts. This is discussed further below. Thus, in another aspect, the invention provides an antibody which is capable of binding a polypeptide or peptide fragment as provided herein. Optionally, the antibody is a monoclonal antibody. A polypeptide or peptide according to the present invention may be isolated and/or purified (e.g. using an antibody) for instance after production by expression from encoding nucleic acid. Polypeptides or peptides according to the present invention may also be generated wholly or partly by chemical synthesis.

The polypeptides provided herein can also be used to identify or design agonists or antagonists of an Spn43A serpin which modulates the interaction between the serpin and a serine protease. This will be described in further detail below.

The products provided herein have several useful applications, as described here and in further detail below. The sequence information provided by the present inventors may be used to introduce genetic defects which result in the loss or reduction of serpin function (eg the loss or reduction of the ability of Spn43Ac to inhibit serine protease activity), generally termed "loss-of-function" serpin mutations herein.

These mutations can be used in a number of ways. As described above, the involvement of the *necrotic* gene (*Spn43Ac*) has been implicated in *Drosophila* immune response pathways. The disclosures of the present inventors and the sequence information herein allow the provision and use of mutant forms of the *necrotic* gene, and/or the peptides encoded thereby, to modulate such pathways. Antibodies and antagonists as further provided by the present inventors

can be used in a similar way, to block the inhibitory effect of serpins on immune response pathways. Furthermore, antisense oligonucleotides of the present invention can be used to inhibit the expression of *Spn43Ac* serpin genes, for the same purpose.

Thus, in one embodiment, the interaction between the Spn43Ac necrotic polypeptide and a serine protease can be diminished by generating a loss-of-function mutation in the *necrotic* gene or polypeptide which affects the serine protease-interacting domain of the serpin. The mutation may be one or more of a deletion (eg as exemplified by transheterozygous *Df(2R)nap-2/Df(2R)sple-D1* flies; see Figure 3), a substitution, an insertion, or any other type of mutation which provides a loss-of-function phenotype in the organism. In another embodiment, the interaction between the Spn43Ac serpin and a serine protease can be reduced by use of an antibody or antagonist of the necrotic polypeptide as provided by the present invention, which competitively or non-competitively inhibits either or both of the serpin or the serine protease. In a yet further embodiment, the level of expression of *Spn43Ac* genes can be reduced by use of an antisense molecule, which binds to a sequence as provided herein. The reduced level of Spn43Ac serpin polypeptides in the cell or medium could then be exploited to the effect of reducing the inhibitory effect of the Spn43Ac serpin on the immune response cascade. Thus the present invention provides methods based on the uses as set out above. By use of such methods, the downstream expression of anti-fungal or anti-viral peptides can thus be stimulated.

The present invention also provides methods which involve obtaining a cDNA library from *nec* mutant fat body cells or a *nec* mutant-derived tissue culture system, in which the

expression of anti-fungal or anti-viral peptides is stimulated as explained above, expressing said cDNA library in a host cell culture, and harvesting the peptides thus expressed. The expressed peptides may then be incorporated into pharmaceutical compositions for use in the treatment of, for instance, fungal or viral infection. The composition may include suitable excipients, carriers and/or adjuvants, as is commonly known in the art.

Reduction of the inhibitory effect of Spn43Ac serpins on serine protease activity can also be employed to identify peptides and/or their encoding nucleic acid sequences, the expression of which is normally controlled by this inhibitory effect. For instance, using such a method, the expression of peptides which are normally upregulated in response to an immune challenge can be achieved. Thus the present invention provides a method of identifying peptides, and/or their encoding nucleic acid sequences, by diminishing the inhibitory effect of Spn43Ac serpin molecules on serine protease activity. The interaction between the Spn43Ac necrotic polypeptide and a serine protease can be diminished by an approach as stated above eg use of a loss-of-function *necrotic* mutation, an agent such as an antibody or antagonist or an antisense molecule as provided herein.

The peptides and/or their encoding nucleic acids thus identified by the methods provided may be screened for anti-fungal or anti-viral peptides. The method may also include the step of obtaining a cDNA library from the cell culture to which the method is applied, expressing the cDNA library in a host cell culture, and harvesting the peptides thus expressed. The peptides may be isolated, cloned, and incorporated into a pharmaceutical composition for use in the treatment of, for instance, fungal or viral infection.

In this way, peptides which would normally only be expressed in response to a particular environmental stimulus, such as an immune challenge, can be identified *in vitro*.

In a further aspect, the invention concerns a method of screening for agents which can affect the inhibition of serine protease activity by a serpin, the method comprising contacting a serpin polypeptide or a peptide fragment as provided herein, which has a biological property of a wild-type serpin protein, with a candidate molecule, and monitoring any change in the inhibition of the serine protease activity. The method can, in this way, be used to identify serpin agonists or antagonists.

In further steps, useful agents identified according to this method can be isolated, cloned and incorporated into a pharmaceutical composition.

Brief Description of the Drawings

These and other aspects of the present invention will now be further described with reference to the accompanying drawings, by way of example and not limitation. Further aspects of the invention will be apparent to those of ordinary skill in the art.

Figure 1 shows a schematic diagram of the Toll signalling pathway.

Figure 2 shows the nucleotide sequences of the *Spn43Aa*, *Spn43Ab* and *Spn43Ac* genes of *Drosophila*.

Figure 3 shows a schematic diagram of the *Spn43A* region of the *Drosophila* chromosome 2.

Figure 4 shows expression of antimicrobial peptide genes in wild-type and *nec* mutant adults. A representative Northern blot of total RNA extracted from control and bacteria-challenged wild-type (Or^R) and *necrotic* (*nec*) loss-

of-function mutant flies is shown.

(a) The blot was hybridized successively with the cDNA probes for drosomycin; diptericin; cecropin A1; metchnikowin and rp49. C, control; 6 h, 6 hours after immune-challenge. The flies used were Oregon-R and the transheterozygous nec^1/nec^2 adults.

(b) Expression of the *drosomycin* and *diptericin* gene in the wild type (Or^R) and necrotic loss-of-function (*nec*) mutants. Lanes Spn-c and Spn-a correspond to the total RNA extracts from the transgenic fly lines bearing *Spn43Ac* and *Spn43Aa* genes, respectively, in the *nec*-deficient background. The flies used were:

Or^R:Oregon-R ; *nec*:nec^1/nec^2; $P\{w^+$ $UAS\text{-}Spn43Ac^+\}/+$; *Spn43Ac*:nec^1/nec^2; $P\{w^+$ $pda\text{-}GAL4\}/P\{w^+$ $UAS\text{-}Spn43Ac^+\}$; *Spn-a*:nec^1/nec^2; $P\{w^+$ $pda\text{-}GAL4\}/$ $P\{w^+$ $UAS\text{-}Spn43Aa^+\}$. C, control; 4h, 4 hours after immune-challenge.

Figure 5 shows transcriptional profiles of the *Spn43Ac* gene.

(a) Northern blot analysis was performed with polyadenylated RNA extracted from adult flies. The blot was hybridized successively with Spn43Ac and rp49 cDNA probes. C, control; 6 h, 6 hours after immune-challenge. Flies used were:

Or^R:$Oregon^R$; Tl^-: Tl^{r632}/Tl^{9QRE} (29°C); Tl^D:$Tl^{10b}/+$; *imd:imd/imd*.

(b) Expression of the *Spn43Ac* gene in fungi-infected wild type adults. Oregon-R flies were anaesthetised and covered with spores of *B. bassiana*. Flies were placed at 29°C and collected after different time intervals: 12 hours (12h), 1 day (1d), 2 days (2d), 3 days (3d) and 4 days (4d). C, control.

Figure 6 shows induction of antimicrobial peptide genes in wild-type and mutant adults.

(a) Expression of the *drosomycin* and *diptericin* genes in

the Toll pathway deficient mutants. 20 μg of total RNA extracted from control and bacteria-challenged wild-type, *nec* single and double mutant flies were blotted and hybridized successively with the cDNA probes indicated to the left of the lanes. C, control; 6 h, 6 hours after bacterial-challenge. The flies used were: Or^R:Oregon-R; *nec*:nec^1/nec^2; *nec*; *pll* : nec^1/nec^2; pll^{078}/pll^{21}; *nec*; *spz*:nec^1/nec^2; spz^{rm7}/spz^{rm7}; *nec*; Tl^-:nec^1/nec^2; Tl^{r632}/Tl^{9QRE} (29°C); *nec*; *snk* : nec^1/nec^2; snk^{073}/snk^{073}; *nec*; *gd*:gd^8/gd^8; nec^1/nec^2.

(b) Expression of the *drosomycin* and *diptericin* genes in wild type, *nec* and *nec;imd* mutant adults. Abbreviations are as in Figure 2a. The flies used were: Or^R:Oregon-R; *nec*:nec^1/nec^2; *nec*, *imd*:nec^1, *imd*/nec^2, *imd*. C, control; 4 h, 4 hours after immune-challenge.

Figure 7 shows detection of the Spn43Ac protein in the hemolymph of wild-type flies by Western blot analysis. Hemolymph was extracted from control and bacteria-challenged Oregon-R (Or^R) flies and from flies carrying a transheterozygous combination of deficiencies, *Df(2R)sple-J1/Df(2R)nap2*, that uncovers the *Spn43Ac* gene (def(*Spn-c*)). C, control; 6 h, 6 hours after bacterial-challenge. 5 μg of total protein were analysed by Western blotting and probed with the antiserum directed against recombinant Spn43Ac protein. Both arrows point to specific protein bands that are present in the hemolymph of wild-type flies and absent from the hemolymph of the negative control flies that lack the *Spn43Ac* gene. The arrow marked 1 points to a constitutive form of Spn43Ac at approximately 62 kDa, while the arrow marked 2 points to an inducible form of Spn43Ac at approximately 52 kDa. The molecular weight markers are indicated to the right of the blot.

Figure 8 shows cleavage of Spaetzle protein in the wild type and *nec* loss-of-function mutants.
40 μg of total protein extracts of adult flies were separated by SDS-PAGE and analysed by Western blotting using the antisera directed against the C-terminal part of the Spaetzle protein. Arrow A indicates signals of approximately 45 kDa detected in wild-type (Or^R) and necrotic loss-of-function mutant *(nec)* flies before bacterial challenge (C). This signal was not detectable in the wild-type flies 1 hour (1h) after bacterial challenge or in the mutants deficient in the *spz* gene (data not shown). Two more signals corresponding to the proteins of 16-18 kDa were detected in wild-type flies after bacterial challenge (Arrows B) and these were also present constitutively in *necrotic* loss-of-function mutants.
Figure 9 shows the predicted amino acid sequences of the Spn43Aa, -Ab and -Ac polypeptides.

Detailed Description

The "*Spn43A* region" refers to the portion of *Drosophila* chromosome 2 containing the *Spn43Aa, -Ab and -Ac* loci.
The "*Spn43Ac* locus" includes the *necrotic* gene (both the coding sequences (exons) and intervening sequences (introns)) and its regulatory elements for controlling transcription and/or translation. The term also covers allelic variations within the locus.
The term "*necrotic* gene" or "*necrotic* allele" includes normal alleles of the *Spn43Ac* gene, as well as variants or derivatives of this sequence and also alleles carrying one or more variations that are linked to a predisposition to give the *necrotic* phenotype. However, for the sake of clarity, where discussion is made of a mutation of the *Spn43Ac* gene which results in a functional deficiency of

the Spn43Ac serpin, or to a functionally-deficient mutant version of the Spn43Ac polypeptide, it is herein referred to as a "loss-of-function *necrotic* mutation" or a "loss-of-function necrotic mutant", respectively. Examples of such mutations are a G466S alteration in nec^2, and deletion of I118 and I119 in nec^1.

The terms "*Spn43Aa* nucleic acid", "*Spn43Ab* nucleic acid" and "*Spn43Ac* nucleic acid" include, respectively, a nucleic acid molecule which has the relevant nucleotide sequence shown in Figure 2, or a variant, derivative or allele of this sequence. The sequence may differ from that shown by a change which is one or more of an addition, insertion, deletion and substitution of one or more nucleotides of the sequence shown. Changes to a nucleotide sequence may result in an amino acid change at the protein level, or not, as determined by the genetic code. If there is such a change, and unless otherwise stated, the encoded serpin retains a biological property of a serpin polypeptide as shown in figure 9; eg common immunoreactivity such as cross reactivity with an antibody, ability to inhibit serine protease activity. Where a nucleic acid mutation diminishes this property, this will be made clear by use of the terms "loss-of-function mutation/mutant", as described above.

Thus, nucleic acid provided by the present invention may comprise a sequence which is different from the sequence shown in Figure 2, and yet which encodes a polypeptide with the same or similar amino acid sequence and an equivalent biological property to a serpin polypeptide as shown in figure 9. On the other hand, and where expressly stated, the nucleic acid may encode a functionally-deficient serpin molecule (as in the case of *necrotic* loss-of-function mutants).

The amino acid sequence of the complete necrotic polypeptide encoded by the *Spn43Ac* nucleic acid shown in Figure 2 consists of 476 residues. Figure 9 shows the amino acid sequences of the full-length wild-type Spn43Aa, -Ab and -Ac polypeptides.

Nucleic acid provided by the present invention comprising a sequence which is different from the sequence shown in Figure 2, and yet which encodes a functionally related polypeptide, may show greater than about 20% homology with the coding sequence shown in Figure 2, greater than about 30% homology, greater than about 40% homology, greater than about 50% homology, greater than about 60% homology, greater than about 70% homology, greater than about 80% homology, greater than about 90% homology or greater than about 95% homology.

In accordance with the present invention, nucleic acids having the appropriate level of sequence homology with the protein encoding region of Figure 2 may be identified by using hybridization and washing conditions of appropriate stringency. For example, hybridizations may be performed, according to the method of Sambrook et al., ("Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989) using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 0.5-1.0% SDS, 100 μg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2X SSC and 1% SDS; (2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 1X SSC and 1% SDS; (4) 2 hours at 42-65°C in 1X SSC and 1% SDS, changing the solution every 30 minutes.

One common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology is (Sambrook et al., 1989):

$T_m = 81.5°C + 16.6Log\ [Na+] + 0.41(\%\ G+C) - 0.63\ (\%\ formamide) - 600/\#bp\ in\ duplex$

As an illustration of the above formula, using [Na+] = [0.368] and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C. Such a sequence would be considered substantially homologous to the nucleic acid sequence of the present invention.

The sequence information provided herein can be utilised in the preparation of Spn43Aa, -Ab or -Ac serpins. As well as full-length serpin molecules, fragments, variants or mutant forms are also obtainable, using standard protein technology.

Polypeptides or peptides may be generated wholly or partly by chemical synthesis. Compounds of the present invention can be readily prepared according to well-established, standard liquid or, preferably, solid-phase peptide synthesis methods, general descriptions of which are broadly available (see, for example, in J.M. Stewart and J.D. Young, Solid Phase Peptide Synthesis, 2nd edition, Pierce Chemical Company, Rockford, Illinois (1984), in M. Bodanzsky and A. Bodanzsky, The Practice of Peptide Synthesis, Springer Verlag, New York (1984); and Applied Biosystems 430A Users Manual, ABI Inc., Foster City, California), or they may be prepared in solution, by the liquid phase method or by any combination of solid-phase,

liquid phase and solution chemistry, e.g. by first completing the respective peptide portion and then, if desired and appropriate, after removal of any protecting groups being present, by introduction of the residue X by reaction of the respective carbonic or sulfonic acid or a reactive derivative thereof.

Another convenient way of producing a peptidyl molecule according to the present invention (peptide or polypeptide) is to express nucleic acid encoding it, by use of nucleic acid in an expression system.

Generally, nucleic acid according to the present invention is provided as an isolate, in isolated and/or purified form, or free or substantially free of material with which it is naturally associated, such as free or substantially free of nucleic acid flanking the gene in the human genome, except possibly one or more regulatory sequence(s) for expression.

Nucleic acid may be wholly or partially synthetic and may include genomic DNA, cDNA or RNA. Where nucleic acid according to the invention includes RNA, reference to the sequence shown should be construed as reference to the RNA equivalent, with U substituted for T.

Nucleic acid sequences encoding a polypeptide or peptide in accordance with the present invention can be readily prepared by the skilled person using the information and references contained herein and techniques known in the art (for example, see Sambrook, Fritsch and Maniatis, "Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, and Ausubel et al, Short Protocols in Molecular Biology, John Wiley and Sons, 1992), given the nucleic acid sequence and clones available. These techniques include (I) the use of the polymerase chain reaction (PCR) to amplify samples of such nucleic acid,

e.g. from genomic sources, (ii) chemical synthesis, or (iii) preparing cDNA sequences. DNA encoding Spn43A serpin fragments may be generated and used in any suitable way known to those of skill in the art, including by taking encoding DNA, identifying suitable restriction enzyme recognition sites either side of the portion to be expressed, and cutting out said portion from the DNA. The portion may then be operably linked to a suitable promoter in a standard commercially available expression system. Another recombinant approach is to amplify the relevant portion of the DNA with suitable PCR primers. Modifications to the serpin sequences can be made, e.g. using site-directed mutagenesis, to lead to the expression of modified serpin peptide or to take account of codon preference in the host cells used to express the nucleic acid.

In order to obtain expression of the nucleic acid sequences, the sequences can be incorporated in a vector having one or more control sequences operably linked to the nucleic acid to control its expression. The vectors may include other sequences such as promoters or enhancers to drive the expression of the inserted nucleic acid, nucleic acid sequences so that the polypeptide or peptide is produced as a fusion and/or nucleic acid encoding secretion signals so that the polypeptide produced in the host cell is secreted from the cell. Polypeptide can then be obtained by transforming the vectors into host cells in which the vector is functional, culturing the host cells so that the polypeptide is produced, and recovering the polypeptide from the host cells or the surrounding medium. Prokaryotic and eukaryotic cells are used for this purpose in the art, including strains of E. coli, yeast, and eukaryotic cells such as COS or CHO cells.

Vectors such as viral vectors have been used in the prior art to introduce nucleic acid into a wide variety of different target cells. Typically the vectors are exposed to the target cells so that transfection can take place in a sufficient proportion of the cells to provide a useful therapeutic or prophylactic effect from the expression of the desired polypeptide. The transfected nucleic acid may be permanently incorporated into the genome of each of the targeted tumour cells, providing long lasting effect, or alternatively the treatment may have to be repeated periodically.

A variety of vectors, both viral vectors and plasmid vectors, are known in the art, see US Patent No. 5,252,479 and WO 93/07282. In particular, a number of viruses have been used as gene transfer vectors, including papovaviruses, such as SV40, vaccinia virus, herpesviruses, including HSV and EBV, and retroviruses. Many gene therapy protocols in the prior art have used disabled murine retroviruses.

As an alternative to the use of viral vectors other known methods of introducing nucleic acid into cells includes electroporation, calcium phosphate co-precipitation, mechanical techniques such as microinjection, transfer mediated by liposomes and direct DNA uptake and receptor-mediated DNA transfer.

Receptor-mediated gene transfer, in which the nucleic acid is linked to a protein ligand via polylysine, with the ligand being specific for a receptor present on the surface of the target cells, is an example of a technique for specifically targeting nucleic acid to particular cells.

Antisense oligonucleotides may be designed to hybridise to the complementary sequence of nucleic acid, pre-mRNA or mature mRNA, for interfering with the production of serpin

polypeptide encoded by a given DNA sequence, so that its expression is reduced or prevented altogether. In addition to the serpin encoding sequence, antisense techniques can be used to target the control sequences of the serpin gene, e.g. in the 5' flanking sequence of the serpin encoding sequence, whereby the antisense oligonucleotides can interfere with serpin control sequences. The construction of antisense sequences and their use is described in Peyman and Ulman, Chemical Reviews, 90:543-584, (1990), Crooke, Ann. Rev. Pharmacol. Toxicol., 32:329-376, (1992), and Zamecnik and Stephenson, P.N.A.S, 75:280-284, (1974).

The present invention also encompasses a method of making a polypeptide or peptide (as disclosed), the method including causing expression from nucleic acid encoding the polypeptide or peptide (generally, nucleic acid according to the invention). This may conveniently be achieved by growing a host cell in culture, containing such a vector, under appropriate conditions which cause or allow expression of the polypeptide. Polypeptides and peptides may also be expressed in *in vitro* systems, such as reticulocyte lysates.

Systems for cloning and expression of a polypeptide in a variety of different host cells are well known. Suitable host cells include bacteria, eukaryotic cells such as mammalian and yeast, and baculovirus systems. Mammalian cell lines available in the art for expression of a heterologous polypeptide include Chinese hamster ovary cells, HeLa cells, baby hamster kidney cells, COS cells and many others. A common, preferred bacterial host is *E. coli*.

Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences including promoter sequences, terminator fragments, polyadenylation sequences,

enhancer sequences, marker genes and other sequences as appropriate. Vectors may be plasmid-derived or viral e.g. 'phage, or phagemid, as appropriate. For further details see, for example, Molecular Cloning: a Laboratory Manual: 2nd edition, Sambrook et al., 1989, Cold Spring Harbor Laboratory Press. Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in Current Protocols in Molecular Biology, Ausubel et al. eds., John Wiley & Sons, 1992.

A further aspect of the present invention provides a host cell containing heterologous nucleic acid, the nucleic acid being as disclosed herein. The nucleic acid of the invention may be integrated into the genome (e.g. chromosome) of the host cell. Integration may be promoted by inclusion of sequences which promote recombination with the genome, in accordance with standard techniques. The nucleic acid may be on an extra-chromosomal vector within the cell, or otherwise identifiably heterologous or foreign to the cell.

A still further aspect provides a method which includes introducing the nucleic acid into a host cell. The introduction, which may (particularly for *in vitro* introduction) be generally referred to without limitation as "transformation", may employ any available technique. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride

transformation, electroporation and transfection using bacteriophage. As an alternative, direct injection of the nucleic acid could be employed.

Marker genes such as antibiotic resistance or sensitivity genes may be used in identifying clones containing nucleic acid of interest, as is well known in the art.

The introduction may be followed by causing or allowing expression from the nucleic acid, e.g. by culturing host cells (which may include cells actually transformed although more likely the cells will be descendants of the transformed cells) under conditions for expression of the gene, so that the encoded polypeptide (or peptide) is produced. If the polypeptide is expressed coupled to an appropriate signal leader peptide it may be secreted from the cell into the culture medium. Following production by expression, a polypeptide or peptide may be isolated and/or purified from the host cell and/or culture medium, as the case may be, and subsequently used as desired, e.g. in the formulation of a composition which may include one or more additional components, such as a pharmaceutical composition which includes one or more pharmaceutically acceptable excipients, vehicles or carriers (e.g. see below).

Thus, a host cell containing nucleic acid according to the present invention, e.g. as a result of introduction of the nucleic acid into the cell or into an ancestor of the cell and/or genetic alteration of the sequence endogenous to the cell or ancestor (which introduction or alteration may take place *in vivo* or *ex vivo*), may be comprised (e.g. in the soma) within an organism which is an animal, particularly a mammal, which may be human or non-human, such as rabbit, guinea pig, rat, mouse or other rodent, cat, dog, pig, sheep, goat, cattle or horse, or which is a bird, such as a chicken. Since it is thought that there is notable

homology between *Drosophila* serpins and serpins from other species (including humans), the products provided by the present invention can be used to investigate corresponding serpins in other organisms. For instance, molecules identified according to the screening methods described herein, which interfere with the interaction between an Spn43Ac serpin and a serine protease, may prove to be useful in interfering with the corresponding interaction between a serpin and serine protease in e.g. a human immune response cascade.

This may have a therapeutic aim. Hence, the presence of a mutant, allele, derivative or variant sequence of e.g an Spn43Ac serpin within cells of an organism, particularly when in place of a homologous endogenous sequence, may allow the organism to be used as a model in testing and/or studying substances which modulate activity of the encoded polypeptide *in vitro* or are otherwise indicated to be of therapeutic potential. Conveniently, assays for such substances may be carried out *in vitro*, within host cells or in cell-free systems.

Suitable screening methods are conventional in the art. They include techniques such as radioimmunosassay, scintillation proximetry assay and ELISA methods. Suitably, either the serpin protein or a serine protease, or a fragment, analogue, derivative, variant or functional mimetic thereof, is immobilised. The other member of the interacting pair is then applied in the presence of the agents under test. In a scintillation proximetry assay, a biotinylated protein fragment is bound to streptavidin-coated scintillant-impregnated beads (produced by Amersham). Binding of radiolabelled peptide is then measured by determination of radioactivity-induced scintillation as the radioactive peptide binds to the

immobilized fragment. Agents which intercept this are thus inhibitors of the interaction.

In one general aspect, the present invention provides an assay method for identifying substances with the ability to alter the inhibitory effect of a serpin on serine protease activity. The method may include:

(a) bringing into contact a substance according to the invention including a peptide fragment of an Spn43A serpin or a derivative, variant or analogue thereof as disclosed, a substance including the relevant fragment of the serine protease or a variant, derivative or analogue thereof, and a test compound, under conditions wherein, in the absence of the test compound which proves to alter interaction between the serpin and the serine protease, said substances interact so that the serine protease activity is inhibited; and,

(b) monitoring any alteration of serine protease inhibition.

A test compound which modulates (eg disrupts, reduces, enhances, interferes with, wholly or partially abolishes) the inhibition of a serine protease by a serpin, may thus be identified.

Performance of an assay method according to the present invention for screening for modulators, may be followed by isolation and/or manufacture and/or use of a compound, substance or molecule which tests positive for ability to interfere with the inhibition by a serpin of serine protease activity.

The precise format of such an assay may be varied by those of skill in the art using routine skill and knowledge. For example, interaction between substances may be studied *in vitro* by labelling one with a detectable label and bringing it into contact with the other which has been immobilised

on a solid support. Suitable detectable labels, especially for peptidyl substances, include ^{35}S-methionine which may be incorporated into recombinantly produced peptides and polypeptides. Recombinantly produced peptides and polypeptides may also be expressed as a fusion protein containing an epitope which can be labelled with an antibody. The protein which is immobilized on a solid support may be immobilized using an antibody against that protein bound to a solid support or via other technologies which are known *per se*. A preferred *in vitro* interaction may utilise a fusion protein including glutathione-S-transferase (GST). This may be immobilized on glutathione agarose beads. In an *in vitro* assay format of the type described above, a test compound can be assayed by determining its ability to diminish the amount of labelled peptide or polypeptide which binds to the immobilized GST-fusion polypeptide. This may be determined by fractionating the glutathione-agarose beads by SDS-polyacrylamide gel electrophoresis. Alternatively, the beads may be rinsed to remove unbound protein and the amount of protein which has bound can be determined by counting the amount of label present in, for example, a suitable scintillation counter. An assay according to the present invention may also take the form of an *in vivo* assay. The *in vivo* assay may be performed in a cell line such as a yeast strain or mammalian cell line in which the relevant polypeptides or peptides are expressed from one or more vectors introduced into the cell.

The ability of a test compound to modulate eg disrupt interaction or binding between a serpin and a serine protease may be determined using a so-called two-hybrid assay. For example, a polypeptide or peptide containing a fragment of serpin or serine protease as the case may be,

or a peptidyl analogue or variant thereof as disclosed, may be fused to a DNA binding domain such as that of the yeast transcription factor GAL 4. The GAL 4 transcription factor includes two functional domains. These domains are the DNA binding domain (GAL4DBD) and the GAL4 transcriptional activation domain (GAL4TAD). By fusing one polypeptide or peptide to one of those domains and another polypeptide or peptide to the respective counterpart, a functional GAL 4 transcription factor is restored only when two polypeptides or peptides of interest interact. Thus, interaction of the polypeptides or peptides may be measured by the use of a reporter gene probably linked to a GAL 4 DNA binding site which is capable of activating transcription of said reporter gene. This assay format is described by Fields and Song, Nature 340;245-246, 1989. This type of assay format can be used in both mammalian cells and in yeast. Other combinations of DNA binding domain and transcriptional activation domain are available in the art and may be preferred, such as the LexA DNA binding domain and the VP60 transcriptional activation domain.

To take a Lex/VP60 two hybrid screen by way of example for the purpose of illustration, yeast or mammalian cells may be transformed with a reporter gene construction which expresses a selective marker protein (e.g. encoding β-galactosidase or luciferase). The promoter of that gene is designed such that it contains binding site for the LexA DNA-binding protein. Gene expression from that plasmid is usually very low. Two more expression vectors may be transformed into the yeast containing the selectable marker expression plasmid, one containing the coding sequence for the full length *LexA* gene linked to a multiple cloning site. This multiple cloning site is used to clone a gene of interest, i.e. encoding a serpin or serine protease

polypeptide or peptide in accordance with the present invention, in-frame, onto the *LexA* coding region. The second expression vector then contains the activation domain of the herpes simplex transactivator VP16 fused to a test peptide sequence or more preferably a library of sequences encoding peptides with diverse, e.g. random sequences. Those two plasmids facilitate expression from the reporter construct containing the selectable marker only when the LexA fusion construct interacts with a polypeptide or peptide sequence derived from the peptide library.

A modification of this when looking for peptides or other substances which interfere with interaction between a serpin polypeptide or peptide and a serine protease polypeptide or peptide, employs the serpin or serine protease polypeptide or peptide as a fusion with the LexA DNA binding domain, and the counterpart serine protease or serpin polypeptide or peptide as a fusion with VP60, and involves a third expression cassette, which may be on a separate expression vector, from which a peptide or a library of peptides of diverse and/or random sequence may be expressed. A reduction in reporter gene expression (e.g. in the case of β-galactosidase a weakening of the blue colour) results from the presence of a peptide which disrupts the serpin-serine protease interaction, which interaction is required for transcriptional activation of the β-galactosidase gene. Where a test substance is not peptidyl and may not be expressed from encoding nucleic acid within a said third expression cassette, a similar system may be employed with the test substance supplied exogenously. As noted, instead of using LexA and VP60, other similar combinations of proteins which together form a functional transcriptional activator may be used, such as

the GAL4 DNA binding domain and the GAL4 transcriptional activation domain.

When performing a two hybrid assay to look for substances which interfere with the interaction between two polypeptides or peptides it may be preferred to use mammalian cells instead of yeast cells. The same principles apply and appropriate methods are well known to those skilled in the art.

The amount of test substance or compound which may be added to an assay of the invention will normally be determined by trial and error depending upon the type of compound used. Typically, from about 0.01 to 100 nM concentrations of putative inhibitor compound may be used, for example from 0.1 to 10 nM. Greater concentrations may be used when a peptide is the test substance.

Compounds for test may be natural or synthetic chemical compounds used in drug screening programmes. Extracts of plants which contain several characterised or uncharacterised components may also be used.

Antibodies directed to the site of interaction in either protein (serpin or serine protease) form a further class of putative inhibitor compounds. Candidate inhibitor antibodies may be characterised and their binding regions determined to provide single chain antibodies and fragments thereof which are responsible for disrupting the interaction.

The provision of the sequences of the Spn43A serpin polypeptides enables the production of antibodies able to bind specifically to them. Accordingly, a further aspect of the present invention provides an antibody able to bind specifically to a polypeptide, the sequence of which is given in Figure 9. Such an antibody may be specific in the sense of being able to distinguish between the polypeptide

to which it is able to bind and other serpin polypeptides for which it has no or substantially no binding affinity (e.g. a binding affinity of about 1000x lower). Specific antibodies bind an epitope on the molecule which is either not present or is not accessible on other molecules. Antibodies according to the present invention may be specific for polypeptide shown in Figure 9, or they may be specific for a particular mutant, variant, allele or derivative polypeptide, as between that molecule and the Figure 9 polypeptide. They may also be cross-reactive with the equivalent wild-type serpin polypeptide. An antibody may therefore be useful in screening methods and also in methods of modulating serine protease inhibition by the serpins, as discussed herein.

Antibodies are also useful in purifying the polypeptide or polypeptides to which they bind, e.g. following production by recombinant expression from encoding nucleic acid. Preferred antibodies according to the invention are isolated, in the sense of being free from contaminants such as antibodies able to bind other polypeptides. Monoclonal antibodies are preferred for some purposes, though polyclonal antibodies are within the scope of the present invention.

Antibodies may be obtained using techniques which are standard in the art. Methods of producing antibodies include immunising a mammal (e.g. mouse, rat, rabbit, horse, goat, sheep or monkey) with the protein or a fragment thereof. Antibodies may be obtained from immunised animals using any of a variety of techniques known in the art, and screened, preferably using binding of antibody to antigen of interest. For instance, Western blotting techniques or immunoprecipitation may be used (Armitage et al, Nature 357:80-82, 1992). Isolation of

antibodies and/or antibody-producing cells from an animal may be accompanied by a step of sacrificing the animal. As an alternative or supplement to immunising a mammal with a peptide, an antibody specific for a protein may be obtained from a recombinantly-produced library of expressed immunoglobulin variable domains, e.g. using lambda bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on their surfaces; for instance see WO92/01047. The library may be naive, that is, constructed from sequences obtained from an organism which has not been immunised with any of the proteins (or fragments), or may be one constructed using sequences obtained from an organism which has been exposed to the antigen of interest.

Antibodies according to the present invention may be modified in a number of ways. Indeed the term "antibody" should be construed as covering any binding substance having a binding domain with the required specificity. Thus the invention covers antibody fragments, derivatives, functional equivalents and homologues of antibodies, including synthetic molecules and molecules the shape of which mimics that of an antibody, enabling it to bind an antigen or epitope. Example antibody fragments, capable of binding an antigen or other binding partner, are the Fab fragment consisting of the V_L, V_H, Cl and CH1 domains; the Fd fragment consisting of the V_H and CH1 domains; the Fv fragment consisting of the V_L and V_H domains of a single arm of an antibody; the dAb fragment which consists of a V_H domain; isolated CDR regions and F(ab')2 fragments, a bivalent fragment including two Fab fragments linked by a disulphide bridge at the hinge region. Single chain Fv fragments are also included.

Hybridomas capable of producing antibody with desired

binding characteristics are within the scope of the present invention, as are host cells, eukaryotic or prokaryotic, containing nucleic acid encoding antibodies (including antibody fragments) and capable of their expression. The invention also provides methods of production of the antibodies including growing a cell capable of producing the antibody under conditions in which the antibody is produced, and preferably secreted.

The production of monoclonal antibodies is well established in the art. A hybridoma producing a monoclonal antibody according to the present invention may be subject to genetic mutation or other changes. It will further be understood by those skilled in the art that a monoclonal antibody can be subjected to the techniques of recombinant DNA technology to produce other antibodies or chimeric molecules which retain the specificity of the original antibody. Such techniques may involve introducing DNA encoding the immunoglobulin variable region, or the complementarity determining regions (CDRs), of an antibody to the constant regions, or constant regions plus framework regions, of a different immunoglobulin. See, for instance, EP184187A, GB 2188638A or EP-A-0239400. Cloning and expression of chimeric antibodies are described in EP-A-0120694 and EP-A-0125023.

The reactivities of antibodies on a sample may be determined by any appropriate means. Tagging with individual reporter molecules is one possibility. The reporter molecules may directly or indirectly generate detectable, and preferably measurable, signals. The linkage of reporter molecules may be directly or indirectly, covalently, e.g. *via* a peptide bond, or non-covalently. Linkage *via* a peptide bond may be as a result of recombinant expression of a gene fusion encoding

antibody and reporter molecule. One favoured mode is by covalent linkage of each antibody with an individual fluorochrome, phosphor or laser dye with spectrally isolated absorption or emission characteristics. Suitable fluorochromes include fluorescein, rhodamine, phycoerythrin and Texas Red. Suitable chromogenic dyes include diaminobenzidine. Other reporters include macromolecular colloidal particles or particulate material such as latex beads that are coloured, magnetic or paramagnetic, and biologically or chemically active agents that can directly or indirectly cause detectable signals to be visually observed, electronically detected or otherwise recorded. These molecules may be enzymes which catalyse reactions that develop or change colours or cause changes in electrical properties, for example. They may be molecularly excitable, such that electronic transitions between energy states result in characteristic spectral absorptions or emissions. They may include chemical entities used in conjunction with biosensors. Biotin/avidin or biotin/streptavidin and alkaline phosphatase detection systems may be employed. The mode of determining binding is not a feature of the present invention and those skilled in the art are able to choose a suitable mode according to their preference and general knowledge.

Antibodies may also be used in purifying and/or isolating a polypeptide or peptide according to the present invention, for instance following production of the polypeptide or peptide by expression from encoding nucleic acid therefor. Antibodies may also be useful to disrupt serpin-serine protease interaction with a view to reducing inhibition of serine proteases. Antibodies can, for instance, be microinjected into cells, e.g. to reduce the

inhibition of serine protease activity and so to stimulate the immune response cascade in an organism.

Antibodies according to the present invention can also be used to identify molecules from other organisms (e.g humans) which are cross-immunoreactive with the molecules against which the antibodies were raised. In this way, for instance, human homologues of *Drosophila* Spn43A serpins may be identified and isolated.

An antibody may be provided in a kit, which may include instructions for use of the antibody, e.g. in determining the presence of a particular substance in a test sample. One or more other reagents may be included, such as labelling molecules, buffer solutions, elutants and so on. Reagents may be provided within containers which protect them from the external environment, such as a sealed vial.

Other candidate inhibitor compounds of serpin-serine protease interactions may be based on modelling the 3-dimensional structure of a polypeptide or peptide fragment as provided herein, and using rational drug design to provide potential inhibitor compounds with particular molecular shape, size and charge characteristics. As noted, the agent may be peptidyl, e.g. a peptide which includes a sequence as recited above, or may be a functional analogue of such a peptide.

As used herein, the expression "functional analogue" relates to peptide variants or organic compounds having an functional activity equatable with the peptide in question, which may interfere with the binding between native serpin and serine protease.

In a further aspect, the present invention provides the use of the substances provided herein in methods of designing or screening for mimetics of the substances.

Accordingly, the present invention provides a method of

designing mimetics of serpins having serine protease-inhibiting activity, said method comprising:

(I) analysing a substance (eg a serpin as provided) having the biological activity to determine the amino acid residues essential and important for the activity to define a pharmacophore; and,

(ii) modelling the pharmacophore to design and/or screen candidate mimetics having the biological activity. Suitable modelling techniques are known in the art. This includes the design of so-called "mimetics" which involves the study of the functional interactions between the molecules and the design of compounds which contain functional groups arranged in such a manner that they could reproduce those interactions.

The designing of mimetics to a known pharmaceutically active compound is a known approach to the development of pharmaceuticals based on a "lead" compound. This might be desirable where the active compound is difficult or expensive to synthesise or where it is unsuitable for a particular method of administration, e.g. peptides are not well suited as active agents for oral compositions as they tend to be quickly degraded by proteases in the alimentary canal. Mimetic design, synthesis and testing may be used to avoid randomly screening large number of molecules for a target property.

There are several steps commonly taken in the design of a mimetic from a compound having a given target property. Firstly, the particular parts of the compound that are critical and/or important in determining the target property are determined. In the case of a peptide, this can be done by systematically varying the amino acid residues in the peptide, e.g. by substituting each residue in turn. These parts or residues constituting the active

region of the compound are known as its "pharmacophore". Once the pharmacophore has been found, its structure is modelled according to its physical properties, e.g. stereochemistry, bonding, size and/or charge, using data from a range of sources, e.g. spectroscopic techniques, X-ray diffraction data and NMR. Computational analysis, similarity mapping (which models the charge and/or volume of a pharmacophore, rather than the bonding between atoms) and other techniques can be used in this modelling process. In a variant of this approach, the three-dimensional structure of the ligand and its binding partner are modelled. This can be especially useful where the ligand and/or binding partner change conformation on binding, allowing the model to take account of this in the design of the mimetic.

A template molecule is then selected onto which chemical groups which mimic the pharmacophore can be grafted. The template molecule and the chemical groups grafted on to it can conveniently be selected so that the mimetic is easy to synthesise, is likely to be pharmacologically acceptable, and does not degrade *in vivo*, while retaining the biological activity of the lead compound. The mimetic or mimetics found by this approach can then be screened to see whether they have the target property, or to what extent they exhibit it. Further optimisation or modification can then be carried out to arrive at one or more final mimetics for further testing or optimisation, e.g. *in vivo* or clinical testing.

The mimetic or mimetics found by this approach can then be screened to see whether they have the target property, or to what extent they exhibit it. Further optimisation or modification can then be carried out to arrive at one or more final mimetics for *in vivo* or clinical testing.

Mimetics of this type together with their use in therapy form a further aspect of the invention.

The present invention further provides the use of a peptide which includes a sequence as disclosed, or a derivative, active portion, analogue, variant or mimetic thereof, which is able to inhibit serine protease activity, in screening for a substance able to interfere with the serine protease inhibitory effect of the serpin.

One class of agents that can be used to disrupt the binding of an Spn43A serpin and a serine protease are peptides based on the sequence motifs of the serpin that interact with serine protease. Such peptides tend to be small molecules, and may be about 40 amino acids in length or less, preferably about 35 amino acids in length or less, more preferably about 30 amino acids in length, or less, more preferably about 25 amino acids or less, more preferably about 20 amino acids or less, more preferably about 15 amino acids or less, more preferably about 10 amino acids or less, or 9, 8, 7, 6 5 or less in length.

The present invention also encompasses peptides which are sequence variants or derivatives of a wild-type serpins sequence.

Preferably, the amino acid sequence shares homology with a fragment of the serpin sequence shown, preferably at least about 20%, or 30%, or 40%, or 50%, or 60%, or 70%, or 75%, or 80%, or 85% homology, or at least about 90% or 95% homology. Thus, a peptide fragment of the serpin may include 1, 2, 3, 4, 5, greater than 5, or greater than 10 amino acid alterations such as substitutions with respect to a serpin polypeptide sequence as shown in figure 9.

A derivative of a peptide for which the specific sequence is disclosed herein may be in certain embodiments the same length or shorter than the specific peptide. In other

embodiments the peptide sequence or a variant thereof may be included in a larger peptide, as discussed above, which may or may not include an additional portion of serpin. 1, 2, 3, 4 or 5 or more additional amino acids, adjacent to the relevant specific peptide fragment in the serpin, or heterologous thereto, may be included at one end or both ends of the peptide.

As is well-understood, homology at the amino acid level is generally in terms of amino acid similarity or identity. Similarity allows for "conservative variation", i.e. substitution of one hydrophobic residue such as isoleucine, valine, leucine or methionine for another, or the substitution of one polar residue for another, such as arginine for lysine, glutamic for aspartic acid, or glutamine for asparagine. Similarity may be as defined and determined by the TBLASTN program, of Altschul et al, J. Mol. Biol., 215:403-10, 1990, which is in standard use in the art. Homology may be over the full-length of the relevant peptide or over a contiguous sequence of about 5, 10, 15, 20, 25, 30 or 35 amino acids, compared with the relevant wild-type amino acid sequence.

As noted, variant peptide sequences and peptide and non-peptide analogues and mimetics may be employed, as discussed further below.

Various aspects of the present invention provide a substance, which may be a single molecule or a composition including two or more components, which comprises a peptide fragment of serpin which comprises a sequence as recited above and/or disclosed elsewhere herein, a peptide consisting essentially of such a sequence, a peptide including a variant, derivative or analogue sequence, or a non-peptide analogue or mimetic which has the ability to inhibit the interaction between a serpin and a serine

protease.

Variants include peptides in which individual amino acids can be substituted by other amino acids which are closely related as is understood in the art and indicated above.

Non-peptide mimetics of peptides are discussed further below.

As noted, a peptide according to the present invention and for use in various aspects of the present invention may include or consist essentially of a fragment of serpin as disclosed, such as a fragment whose sequence is given above. Where one or more additional amino acids are included, such amino acids may be from the serpin or may be heterologous or foreign to the serpin. A peptide may also be included within a larger fusion protein, particularly where the peptide is fused to a non-serpin (i.e. heterologous or foreign) sequence, such as a polypeptide or protein domain.

The invention also includes derivatives of the peptides described above, including the peptide linked to a coupling partner, e.g. an effector molecule, a label, a drug, a toxin and/or a carrier or transport molecule. Techniques for coupling the peptides of the invention to both peptidyl and non-peptidyl coupling partners are well known in the art. In one embodiment, the carrier molecule is a 16 aa peptide sequence derived from the homeodomain of Antennapedia (e.g. as sold under the name "Penetratin"), which can be coupled to a peptide via a terminal Cys residue. The "Penetratin" molecule and its properties are described in WO 91/18981.

Generally, a modulator according to the present invention is provided in an isolated and/or purified form, i.e. substantially pure. This may include being in a composition where it represents at least about 90% active

ingredient, more preferably at least about 95%, more preferably at least about 98%. Such a composition may, however, include inert carrier materials or other pharmaceutically and physiologically acceptable excipients. As noted below, a composition according to the present invention may include, in addition to a modulator compound as disclosed, one or more other molecules of therapeutic use.

The present invention extends in various aspects not only to a substance identified as a modulator of serpin inhibition of serine protease activity in accordance with that disclosed herein, but also a pharmaceutical composition, medicament, drug or other composition comprising such a modulator, and a method of making a pharmaceutical composition comprising admixing such a modulator with a pharmaceutically acceptable excipient, vehicle or carrier, and optionally other ingredients.

A modulator which could be a polypeptide or peptide or other substance or a nucleic acid molecule encoding a polypeptide/peptide modulator, may be provided in a kit, e.g. sealed in a suitable container which protects its contents from the external environment. Such a kit may include instructions for use.

Preparation of *necrotic* loss-of-function nucleic acids, and vectors and host cells incorporating the nucleic acid

The loss-of-function mutant form of the *necrotic* gene (*Spn43Ac*) may, in different embodiments, contain one or more insertions, deletions, substitutions and/or additions of one or more nucleotides compared with the wild-type sequence (such as is shown in Figure 2) which disrupts the gene function or results in an encoded serpin with an impaired ability to inhibit serine protease activity. For instance, a mutation may result in a frame-shift or stop

codon, which affects the nature of the polypeptide produced (if any), or it may comprise a point mutation or gross mutational change to the encoded polypeptide, including insertion, deletion, substitution and/or addition of one or more amino acids or regions in the polypeptide. A mutation in a promoter sequence or other regulatory region may prevent or reduce expression from the gene or affect the processing or stability of the mRNA transcript.

Particular deletion mutations of the *Spn43Ac* gene used in the experiments of the present inventors are set out in Figure 3. These mutations are generally associated with the production of overlapping deletion mutants from which have been removed all three Spn43A transcripts, *Df(2R)sple-D1/Df(2R)nap-2*, or the two most distal transcripts (*Spn43Ab* and *-Ac*), *Df(2R)sple-D2/Df(2R)nap-2*. These deletion mutants have amorphic *pk* and *nec* mutant phenotypes, but are otherwise wild-type.

Screening for the presence of one or more of these in a test sample has a potential use in, for instance, the detection of anti-fungal or anti-viral peptides, as discussed herein.

A cell containing mutated nucleic acid according to the present invention, e.g. as a result of introducing the mutated nucleic acid into the cell, or into an ancestor of the cell, and/or genetic alteration of the sequence endogenous to the cell or ancestor (which introduction or alteration may take place *in vivo* or *ex vivo*), may, in one embodiment, be comprised within the fat bodies of *Drosophila* flies, preferably adult flies. Also suitable are haemocytes or epithelial cells.

Cells which possess a *necrotic* loss-of-function mutation, and which therefore have a deficiency of wild-type Spn43Ac serpins that would normally inhibit serine proteases of the

immune response cascade, can be used to produce cDNA libraries. These libraries, which will be enriched in nucleic acids normally expressed at high levels only in response to an immune challenge, can then be used to transform suitable host cells, as is known in the art. The host cells can then be cultured under conditions suitable for expression of the encoded peptides. Given the extent to which the normal activity of loss-of-function *necrotic* cells is suppressed, such that the immune response pathway is freed from the inhibitory effects of Spn43Ac serpins, any clone from the mutant cells' cDNA library might be a potential anti-fungal or anti-viral agent. Hence, the proteins thus expressed can be harvested, cloned and screened for e.g. anti-fungal or anti-viral activity.

Migration shift assays to characterise necrotic mutations.

DNA amplification in the PCR.

25ng of genomic DNA from each individual to be screened for mutations was amplified in 35 cycles of the PCR using oligonucleotide primers designed using the sequence information provided in Figure 2. Prior to incorporation into the PCR, both oligonucleotide primers were end-radiolabelled with gamma ^{32}P using T4 polynucleotide kinase. Following amplification in the PCR, formamide loading dye was added to each sample and the sample denatured at 94°C for 3 minutes. Following denaturation the sample was placed immediately on ice.

DNA fragment sizing.

2μl of each sample was loaded immediately onto a well formed by a 40 slot sharks' tooth comb in conventional 0.4mm thick denaturing 6% polyacrylamide gel. The sample was electrophoresed through the gel for 2-5 hours at 90 Watts at room temperature.

SSCP heteroduplex analysis

SSCP is a PCR-based assay for screening DNA fragments for sequence variants/mutations. It involves amplifying radiolabelled 100-300 bp fragments of the *necrotic* gene, diluting these products and denaturing at 95°C.
The fragments were quick-cooled on ice so that the DNA remained in single-stranded form. These single-stranded fragments of *necrotic* were run through acrylamide-based gels. Differences in the sequence composition caused the single-stranded molecules to adopt different conformations in this gel matrix, making their mobility different from wild-type fragments, thus allowing detecting of mutations in the fragments being analysed relative to a control fragment upon exposure of the gel to X-ray film.
These fragments with altered mobility/conformations were directly excised from the gel and directly sequenced for the mutation. Following denaturation, the sample was cooled on ice for 10 minutes to allow the heteroduplex to form. Each sample was electrophoresed through two different types of gel.
A typical set of conditions for SSCP analysis are as follows: 3μl are electrophoresed overnight at 4 Watts at room temperature through a 6% non denaturing polyacrylamide gel containing 10% glycerol.
3μl are electrophoresed for four hours at 30 Watts in a 4°C cold room through a 4.5% non denaturing polyacrylamide acrylamide gel without glycerol.
Following electrophoresis, gels were dried onto Whatman 3MM paper, and placed in an autoradiography cassette at room temperature for a period ranging from two hours to several days.
Following development of the autoradiograph band shifts in individual samples were detected by eye.

Sequencing of PCR product.

Where a band shift is seen in SSCP heteroduplex or DNA fragment sizing gels, the fragment concerned can be reamplified from the relevant stock genomic DNA and directly sequenced. To sequence PCR product, the product was precipitated with isopropanol, resuspended and sequenced using TaqFS+ Dye terminator sequencing kit. Extension products were electrophoresed on an ABI 377 DNA sequencer and data analysed using Sequence Navigator software.

***Necrotic* PTT Assay**

PTT is another PCR-based screening assay. Fragments of nucleic acid were amplified with primers that contain the consensus Kozak initiation sequences and a T7 RNA polymerase promoter. These extra sequences were incorporated into the 5' primer such that they were in-frame with the native coding sequence of the fragment being analysed. These PCR products were introduced into a coupled transcription/translation system. This reaction allowed the production of *necrotic* RNA from the fragment and translation of this RNA into a necrotic protein fragment. PCR products from controls encoded a protein product of a wild-type size relative to the size of the fragment being analysed. If the PCR product analysed had a frame-shift or nonsense mutation, the assay would yield a truncated protein product relative to controls. The size of the truncated product is related to the position of the mutation.

The relative region of the *necrotic* gene from this sample was sequenced to identify the truncating mutation.

Cells harbouring *necrotic* loss-of-function mutations can be detected by standard techniques. These may include, but are not limited to:

(a) comparing the sequence of nucleic acid in the

sample with the wild-type *necrotic* nucleic acid sequence (e.g. that shown in Figure 2) to determine whether the nucleic acid in the sample contains mutations in respect of the *necrotic* gene; or,

(b) determining the presence in the sample of polypeptide encoded by the *necrotic* gene, and, if present, determining whether the Spn43Ac polypeptide is full length, and/or is mutated, and/or is expressed at the normal level; or,

(c) using DNA fingerprinting to compare the restriction pattern produced when a restriction enzyme cuts nucleic acid from the sample with that of a wild-type *necrotic* sample or from a sample having known mutations in the *necrotic* gene; or,

(d) using a specific binding member capable of binding to a *necrotic* nucleic acid sequence (either a normal sequence or a known mutated sequence) or to the polypeptide encoded thereby, the specific binding member e.g. comprising nucleic acid which is hybridisable with the sequence shown in Figure 2 or an allele thereof, or a substance comprising an antibody domain with specificity for a native or mutated *necrotic* nucleic acid sequence or the polypeptide encoded by it, the specific binding member being labelled so that binding of the specific binding member to its binding partner is detectable;

(e) using PCR involving one or more primers based on a normal or mutated *necrotic* gene sequence, e.g as shown in Figure 2, to screen for normal or mutant *necrotic* gene in a sample from *Drosophila*; or

(f) comparing levels of peptides such as anti-fungal peptides in the cells with those of non-mutant cells.

The presence of a *necrotic* loss-of-function mutation may also allow the mutants to be used as models in testing

and/or studying the role of genes or their encoded proteins which lie either upstream or downstream of *necrotic* in pathways such as that involved in generating an immune response in *Drosophila*. The mutants may also be utilised in studies of substances which modulate the activity of such genes or proteins. These studies may have therapeutic implications, as would be obvious to one skilled in the art.

Preparation of cDNA libraries from *necrotic* loss-of-function mutants.

cDNA libraries from cells with loss-of-function mutant necrotic polypeptides can readily be prepared by the skilled person using the information and references contained herein and techniques known in the art (for example, see Sambrook, Fritsch and Maniatis, "Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, and Ausubel et al, Short Protocols in Molecular Biology, John Wiley and Sons, 1992). These techniques include (I) the isolation, purification and fractionation of mRNA molecules from these cells; (ii) the synthesis of cDNA from these mRNA molecules, and (iii) the construction of vectors for cloning the cDNA molecules. Accordingly, nucleic acid encoding peptides such as anti-fungal or anti-viral peptides may be obtained in the form of a cDNA library prepared from cells with loss-of-function mutant necrotic serpins. The cDNA library thus produced can be screened with probes (such as antibodies to known anti-fungal or anti-viral nucleic acids or their encoded peptides, or oligonucleotides of at least about 20-80 bases which can hybridise to known anti-fungal or anti-viral peptide-encoding nucleic acids) designed to identify genes of interest or the proteins encoded by them. Screening of a cDNA library with a selected probe may be conducted using

standard procedures, such as those described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means of isolating genes such as those encoding anti-fungal peptides from the cDNA library is to use PCR methodology [Sambrook *et al.*, supra: Dieffenbach *et al.*, PCR Primer Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

Where oligonucleotide sequences are used as probes to screen for nucleic acids encoding e.g. anti-fungal peptides, these sequences should be selected so that they are of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide probe is preferably labelled such that it can be detected upon hybridization to nucleic acid in the library being screened. Methods of labelling are well known in the art, and include the use of radiolabels such as ^{32}P-labelled ATP, biotinylation or enzyme labelling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook *et al.*, *supra*.

Where the peptides of interest are those with anti-fungal activity, host cells such as *E. coli* cells expressing the cDNA library can be screened for colonies that suppress the growth of a lawn of fungal cells, such as aspergillus cells.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined through sequence alignment using computer software programs such as BLAST, BLAST2, ALIGN, DNAstar,

and INHERIT which employ various algorithms to measure homology.

As is standard procedure in the preparation of cDNA libraries, nucleic acid molecules prepared from *necrotic* loss-of-function mutant cells provided by the present invention can be incorporated into a vector having control sequences operably linked to the nucleic acids to control their expression. The vectors may include other sequences such as promoters or enhancers to drive the expression of the inserted nucleic acid sequences so that the encoded peptides are produced as fusion proteins, and/or nucleic acid sequences encoding secretion signals, such that the peptides produced in the host cell are secreted from the cell. Peptides encoded by the cDNA library can then be obtained by transforming the vectors into host cells in which the vector is functional, culturing the host cells so that the peptides are produced, and recovering the peptides from the host cells or the surrounding medium. Prokaryotic and eukaryotic cells are used for this purpose in the art, including strains of *E. coli*, yeast, and eukaryotic cells such as COS or CHO cells. The choice of host cell can be used to control the properties of the peptides expressed in those cells, e.g. controlling where the peptides are deposited in the host cells or affecting properties such as its glycosylation.

PCR techniques for the amplification of nucleic acid are described in US Patent No. 4,683,195. In general, such techniques require that sequence information from the ends of the target sequence is known to allow suitable forward and reverse oligonucleotide primers to be designed which are identical or similar to the polynucleotide sequence that is the target for the amplification. PCR comprises the steps of denaturation of the template nucleic acid (if

double-stranded), annealing of primer to target, and polymerisation. The nucleic acid probed or used as template in the amplification reaction may be genomic DNA, cDNA or RNA. PCR can be used to amplify specific sequences from genomic DNA, specific RNA sequences and cDNA transcribed from mRNA, bacteriophage or plasmid sequences. References for the general use of PCR techniques include Mullis *et al.*, Cold Spring Harbor Symp. Quant. Biol., 51:263, (1987), Ehrlich (ed.), PCR technology, Stockton Press, NY, 1989, Ehrlich et al, Science, 252:1643-1650, (1991), "PCR protocols; A Guide to Methods and Applications", Eds. Innis *et al.*, Academic Press, New York, (1990).

Screening cDNA libraries by hybridisation.

Fragments of cDNA obtained by the above methods can be ^{32}P labelled and hybridised to various widely available plated or gridded cDNA libraries. Positive clones can then be isolated, and subject to replating and rehybridisation if necessary until a pure clone has been isolated. DNA can then be made from pure clones and will be sequenced by conventional Sanger dideoxy sequencing on a ABI 377 DNA sequencer.

Screening cDNA libraries by PCR amplification.

Oligonucleotides based on sequences within the peptide sequences identified by the methods described herein can be used in conjunction with oligonucleotides designed to prime from the cloning vector in PCR amplifications of aliquots of widely available cDNA libraries. This will allow amplification of fragments of the cDNA positioned between the currently known fragment and the cloning insertion site. Products of the PCR amplification can then be sequenced using Sanger dideoxy sequencing on an ABI 377 sequencer.

Rapid amplification of cDNA ends (RACE).

Primary cDNAs synthesised from necrotic loss-of-function mutant cell RNAs can be ligated to an oligonucleotide linker. After purification, PCR amplifications can be performed using an oligonucleotide that primes from the cDNA sequence and a second oligonucleotide that primes from the linker. Amplification products will be directly sequenced using Sanger dideoxy sequencing.

The new sequences can then be integrated into the full sequence of the gene by detection of overlaps with previously known components of the sequence.

The screening of cDNA or genomic libraries with selected probes can be conducted using standard procedures, for instance as described in "Short Protocols in Molecular Biology", 2nd edition, John Wiley and Sons (1992) or in "Molecular Cloning", Sambrook, Fritsch and Maniatis, Cold Spring Harbour, (1989).

These techniques allow the full coding sequence of the peptide-encoding gene to be isolated. The full length sequence is defined as the sequence between a translation initiation codon (ATG) and a translation termination codon (TAA, TAG, TGA) between which there is an open reading frame. This in turn can be used to define the intron-exon structure of the gene. Primers can then be designed to flank each exon so that the whole coding sequence of the gene can be amplified from genomic DNA.

The skilled person can use the techniques described herein and others well known in the art to produce large amounts of the peptides identified according to the methods of the present invention, or fragments or active portions thereof, for use as pharmaceuticals, in the development of drugs and for further study into their properties and roles *in vivo*. Host cells may be transfected or transformed with

expression or cloning vectors containing the cDNA libraries thus obtained. The host cells can be cultured in conventional nutrient media modified as appropriate to induce promoters, selecting transformants, or amplifying the genes encoding peptides likely to be of interest, such as anti-fungal peptides. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach. M. Butler, ed. JRL Press, 1991) and Sambrook *et al., supra*. Methods of transfection are known to the ordinarily skilled artisan, for example, methods using $CaPO_4$ and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook *et al., supra*, or electroporation is generally used for prokaryotes or other cells that contain substantial cell wall barriers. Infection with *Agrobacterium tumefaciens* can be used for transformation of certain plant cells, as described by Shaw *et al.*, Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in US Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen *et al.*, J. Bact., 130:946 (1977) and Hsiao *et al.*, Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial

protoplast fusion with intact cells, or use of polycations, e.g, polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown *et al.*, Methods in Enzymology, 185:527 537 (1990) and Mansour *et al.*, Nature 336:348-352 (1988).

In this way, host cells expressing a cDNA library as obtained according to the present invention may be used as nucleic acid factories to replicate nucleic acid of interest, such as that encoding an anti-fungal peptide, in order to generate large amounts of it. Multiple copies of nucleic acid of interest may be made within a cell when coupled to an amplifiable gene such as DHFR. Host cells transformed with nucleic acid of interest, or which are descended from host cells into which nucleic acid was introduced, may be cultured under suitable conditions, e.g. in a fermenter, taken from the culture and subjected to processing to purify the nucleic acid. Following purification, the nucleic acid or one or more fragments thereof may be used as desired, for instance to synthesise the encoded peptides in an expression system.

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, *Enterobacteriaceae* such as *E. coli*. Various *E. coli* strains are publicly available, such as E. *coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635).

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for cDNA-carrying vectors. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host

microorganism.

Suitable host cells for the expression of glycosylated peptides such as anti-fungal or anti-viral peptides are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as *Drosophila* S2 and *Spodoptera* Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham *et al*., J. GenVirol 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumour (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

Nucleic acids of interest as obtained by methods according to the present invention may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors

containing one or more of these components employs standard ligation techniques which are known to the skilled artisan. The peptides, such as anti-fungal peptides, may be produced recombinantly not only directly, but also as fusion polypeptides with one or more heterologous polypeptides, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the anti-fungal peptide-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, e.g., the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α-factor leaders, the latter described in U S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 (published 15 November 1990). In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in

mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins e.g, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the nucleic acid of interest, such as DHFR or thymidine kinase. An appropriate host cell, when wild-type DHFR is employed, is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub *et al.*, Proc. Natl. Acad. Sci. USA 77:4216 (1980).

A suitable selection gene for use in yeast is the *trp*1 gene present in the yeast plasmid YRp7 [Stinchcomb *et al.*, Nature, 282:39 (1979); Kingsman *et al.*, Gene, 7:141 (1979); Tschemper *et al.*, Gene, 10:157 (1980)]. The *trp*1 gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC: No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the nucleic acid sequence of interest to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the β-lactamase and lactose promoter systems [Chang *et al.*, Nature, 275:615 (1978); Goeddel *et al.*, Nature. 281:544 (1979)], alkaline phosphatase, a tryptophan (*trp*) promoter system [Goeddel, Nucleic Acids Res. 8:4057 (1980); EP 36,776], and hybrid promoters such as the *tac* promoter [deBoer *et al.*, Proc.

Natl. Acad. Sci. USA. 80:21-25 (1983)]. Promoters for use in bacterial systems will also contain a Shine-Dalgarno (S.D.) sequence operably linked to the nucleic acid of interest.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman *et al.*, J. Biol. Chem. 255:2073 (1980)] or other glycolytic enzymes [Hess *et al.*, J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry. 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

Transcription of nucleic acid encoding peptides such as anti-fungal or anti-viral peptides from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, e.g., the actin promoter or an immunoglobulin promoter, and from heat-shock

promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding a peptide of interest, such as an anti-fungal or anti-viral peptide, by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin α-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the peptide-encoding sequence, but is preferably located at a site 5' to the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5', and occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding the anti-fungal peptide.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of peptides such as anti-fungal peptides in recombinant vertebrate cell culture are described in Gething *et al.*, Nature, 293:620-695 (1981); Mantei *et al.*, Nature, 281:40-46 (1979); EP 117,060; and EP

117,058.

Genes that are expressed at high levels in *necrotic* loss-of-function mutant cells can be identified by comparison with the pattern of expression of genes in "normal" cells (i.e. having a wild-type Spn43Ac background). As described above, mRNA can be extracted from *nec* mutant and wild-type cells. A representative cDNA library can then be synthesised, amplified, (e.g. by PCR) and displayed (e.g. by DD-PCR) using standard molecular technology. Transcripts identified as being specifically upregulated in mutant cells can then be analysed and tested for biological properties such as anti-fungal or anti-viral properties.

Gene amplification and/or expression of peptides of interest as detected by the methods described herein may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl Acad Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labelled. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labelled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either

monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a known native-sequence peptide, such as a known anti-fungal peptide, or against a synthetic peptide, e.g one based on a known anti-fungal peptide.

Forms of peptides of interest, such as anti-fungal or anti-viral peptides, may be recovered from culture medium or from host cell lysates. If membrane-bound, a peptide of interest can be released from the membrane using a suitable detergent solution (e.g Triton-X 100) or by enzymatic cleavage. Cells employed in expression of peptides obtained by the methods provided by the present invention can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify peptides such as anti-fungal or anti-viral peptides from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants; and metal chelating columns to bind epitope-tagged forms of the peptide of interest. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology 182 (1990); Scopes, Protein Purification: Principles and Practice. Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular peptide to be produced.

Nucleotide sequences (or their complement) encoding peptides obtained according to methods of the present invention, such as anti-fungal or anti-viral peptides, have various applications in the art of molecular biology, including uses for the preparation of the said peptides by the recombinant techniques described herein.

The full-length native sequence peptide gene as detected using the methods described herein, or portions thereof, may be used as hybridization probes for a cDNA library e.g. to isolate homologues of the full-length peptide gene encoding naturally-occurring variants of the peptide of interest from other species. Optionally, the length of the probes will be about 20 to about 50 bases. The hybridization probes may be derived from the nucleotide sequence encoding the peptide of interest, or from genomic sequences including promoters, enhancer elements and introns of native sequence peptide. By way of example, a screening method will comprise isolating the coding region of a detected anti-fungal peptide gene using the known DNA sequence to synthesize a selected probe of about 40 bases. Hybridization probes may be labelled by a variety of labels, including radionucleotides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems. Labelled probes having a sequence complementary to that of the gene encoding a peptide of interest, as identified by a method of the present invention, can be used to screen libraries of human cDNA, genomic DNA or mRNA to determine which members of such libraries the probe hybridizes to.

The probes may also be employed in PCR techniques to generate a pool of sequences for identification of closely related peptide-encoding (e.g anti-fungal peptide-encoding) sequences.

Nucleotide sequences encoding a peptide of interest can also be used to construct hybridization probes for mapping the gene which encodes that peptide. The nucleotide sequences identified by the methods described herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

When the coding sequences for the peptide of interest encode a protein which binds to another protein (e.g. where the peptide is a receptor), they can be used in assays to identify the other proteins, polysaccharides or other molecules involved in the binding interaction. By such methods, inhibitors of the receptor/ligand binding interaction can be identified. This may be of interest if the peptide of interest is involved in an anti-fungal or anti-viral response in conjunction with a binding partner with which it interacts. Proteins involved in such binding interactions can also be used to screen for peptide- or small molecule-inhibitors or agonists of the binding interaction. Also, the peptide of interest can be used to isolate correlative ligand(s). Screening assays can be designed to find lead compounds that mimic the biological activity of a native peptide, such as an anti-fungal or anti-viral peptide, or a receptor for such a peptide. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

The peptides, especially anti-fungal and/or anti-viral peptides, or cDNA libraries enriched with nucleic acids encoding such peptides, as obtained by a method of the present invention, can be formulated according to known methods to prepare pharmaceutically-useful compositions, whereby the peptide product or nucleic acid hereof is combined in admixture with a pharmaceutically acceptable carrier vehicle. Therapeutic formulations are prepared for storage by mixing the active ingredient having the desired degree of purity with optional physiologically acceptable carriers, excipients or stabilizers (Remington's Pharmaceutical Sciences 16th edition, Osol, A. Ed. (1980)), in the form of lyophilized formulations or aqueous solutions. Acceptable carriers, excipients or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone, amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as Tween, Pluronics or PEG.

The formulations to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes, prior to or following lyophilization and reconstitution.

Therapeutic compositions herein generally are placed into a container having a sterile access port, for example, an

intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

The route of administration is in accord with known methods, e.g. injection or infusion by intravenous, intraperitoneal, intracerebral, intramuscular, intraocular, intraarterial or intralesional routes, topical administration, or by sustained release systems.

Dosages and desired drug concentrations of the pharmaceutical compositions may vary depending on the particular use envisioned. The determination of the appropriate dosage or route of administration is well within the skill of an ordinary physician. Animal experiments provide reliable guidance for the determination of effective doses for human therapy. Interspecies scaling of effective doses can be performed following the principles laid down by Mordenti, J. and Chappell, W. "The use of interspecies scaling in toxicokinetics" In Toxicokinetics and New Drug Development, Yacobi *et al.*, Eds., Pergamon Press, New York 1989, pp. 42-96.

Whether it is a peptide (e.g. an anti-fungal or anti-viral peptide), nucleic acid molecule, cDNA library, small molecule or other pharmaceutically useful compound obtained according to the methods of the present invention, that is to be given to an individual, administration is preferably in a "prophylactically effective amount" or a "therapeutically effective amount" (as the case may be, although prophylaxis may be considered therapy), this being sufficient to show benefit to an individual. The patient may or may not be human. The actual amount administered, and rate and time-course of administration, will depend on the nature and severity of what is being treated. Prescription of treatment, e.g. decisions on dosage etc, is within the responsibility of general practitioners and

other medical doctors, and typically takes account of the disorder to be treated, the condition of the individual patient, the site of delivery, the method of administration and other factors known to practitioners. Examples of the techniques and protocols mentioned above can be found in Remington's Pharmaceutical Sciences, 16th edition, Osol, A. (ed), 1980.

Alternatively, targeting therapies may be used to deliver the active agent more specifically to certain types of cell, by the use of targeting systems such as antibody or cell specific ligands. Targeting may be desirable for a variety of reasons; for example if the agent is unacceptably toxic, or if it would otherwise require too high a dosage, or if it would not otherwise be able to enter the target cells.

Instead of administering these agents directly, they could be produced in the target cells by expression from an encoding gene or cDNA library introduced into the cells, e.g. in a viral vector (a variant of the VDEPT technique - see below). The vector could be targeted to the specific cells to be treated, or it could contain regulatory elements which are switched on more or less selectively by the target cells.

Alternatively, the agent could be administered in a precursor form, for conversion to the active form by an activating agent produced in, or targeted to, the cells to be treated. This type of approach is sometimes known as ADEPT or VDEPT; the former involves targeting the activating agent to the cells by conjugation to a cell-specific antibody, while the latter involves producing the activating agent, e.g. an enzyme, in a vector by expression from encoding DNA in a viral vector (see for example, EP-A-415731 and WO 90/07936).

A composition may be administered alone or in combination with other treatments, either simultaneously or sequentially dependent upon the condition to be treated.

Methods involving use of *necrotic* loss-of-function mutants in diagnostic and/or prognostic contexts, for instance in treating fungal or viral infection, are also provided herein.

Nucleic acid according to the present invention, such as a full-length coding sequence or oligonucleotide probe or primer, may be provided as part of a kit, e.g. in a suitable container such as a vial in which the contents are protected from the external environment. The kit may include instructions for use of the nucleic acid, e.g. in PCR and/or a method for determining the presence of nucleic acid of interest in a test sample. A kit wherein the nucleic acid is intended for use in PCR may include one or more other reagents required for the reaction, such as polymerase, nucleosides, buffer solution etc. The nucleic acid may be labelled. A kit for use in determining the presence or absence of nucleic acid of interest may include one or more articles and/or reagents for performance of the method, such as means for providing the test sample itself, e.g. a swab for removing cells from the buccal cavity or a syringe for removing a blood sample (such components generally being sterile). In a further aspect, the present invention provides an apparatus for screening particular nucleic acid, the apparatus comprising storage means including the relevant nucleic acid sequence as provided herein, the stored sequence being used to compare the sequence of the test nucleic acid to determine the presence of mutations.

<u>**Materials and methods**</u>

Drosophila stocks:

Oregon R (Or^R) flies were chosen as a standard wild-type strain. Stocks and crosses were raised on a cornmeal *Drosophila* medium and grown at room temperature (24-25°C) unless otherwise noted. The two *necrotic* (*nec*) alleles (*nec*^{1} and *nec*^{2}) as well as the *Df(2R)sple-J1* and *Df(2R)nap2* deficiencies are described in Heitzler et al., Genetics 135: 105-115 and were obtained from J. Roote (Cambridge, UK). The *Df(2R)sple-J1* and *Df(2R)nap2* deficiencies have been mapped respectively to 43A1; 43C3-C7 and 41F4-F9; 43A1. The mutant lines used in the double mutant analysis have been described elsewhere [Lemaitre,et al., Cell 86, 973 (1996)]. In order to generate the *nec, imd* double mutant, *imd* was recombined with both *nec*^{1} and *nec*^{2} alleles. The transgenic strain carrying a *drosomycin* promoter-GFP reporter gene has been described elsewhere, and was crossed into a *nec*^{1}/*nec*^{2} context. The transgenic strain pAct-GFP is a w^- line carrying a GFP reporter gene under the control of the actin 5C promoter on the *CyO* balancer chromosome (Reichhart J.-M. and Ferrandon D., in press). The stock is available from the Bloomington Stock Center (Bloomington, IN). A strong fluorescence in the salivary duct, the copper cells, the proventriculus and in the visceral musculature of the midgut of larvae renders such a GFP-labelled *CyO* balancer an easy tool to score under the fluorescent dissecting microscope (Reichhart J.-M. and Ferrandon D., in press).

RNA analysis:

Total RNA was extracted by the Trizol reagent (Gibco BRL) according to the supplier's protocol. Polyadenylated RNA was isolated from total RNA by two sequential chromatographies on oligo (dT) cellulose (type 7, Pharmacia). Northern blotting experiments were performed as in Lemaitre *et al.* (EMBO J., 1995, 14(3):536). The

following probes were used: *cecropin A1* cDNA (Kylsten, 1990, EMBO J., 1990,9:217) , *diptericin* cDNA (Wicker, 1990, 265(36):22493), *drosomycin* cDNA (Fehlbaum, 1994, 269(52):33159), *metchnikowin* cDNA (Levashina, 1995, Eur. J. Biochem. 233:694), *Spn43Ac* cDNA (an *EcoRI* fragment of approximately 750 bp corresponding to the 3' region of *Spn43Ac* cDNA; Green *et al.*, in preparation) and *rp49* cDNA (a PCR fragment of approximately 400 bp generated between two oligonucleotides designed after the *rp49* coding sequence; (O'Connell, 1984, Nucl. Acids Res. 12, 5495-5513). The *cecropin A1* probe cross reacts with *cecropin A2* transcripts [Kylsten, 1990 , *supra*).

Antibody production and Western blot analysis :

A chimeric protein composed of Glutathione S-Transferase (GST) fused to the *Spn43Ac* gene product was produced using a GST-SPN43Ac expression vector: a *SacI* (filled in) -*XhoI* 1200 bp fragment of *Spn43Ac* cDNA was subcloned into the *SmaI* and *XhoI* sites of the pGEX2T expression vector (Pharmacia). The GST-Spn43Ac fusion protein was expressed in the *E. coli* strain LE 392. One litre of bacterial culture was grown to an O.D. of 0.5 at 37°C. After induction with IPTG (0.4 mM) and 6 hours of culture at 30°C, cells were pelleted by centrifugation, washed with PBS (Phosphate Buffered Saline) and resuspended in 30 ml of lysis buffer (PBS; 0.1 mM PMSF; 1% Triton; protease inhibitors). Bacteria were sonicated (1 min; 10 times) and centrifuged at 12000 g for 10 min at 4°C. The recombinant protein was localized in the inclusion bodies and these were extracted from the pellet using a Sarkosyl based method as described in Grieco, 1992 (Nucl. Acid Res. 20(24):6733). In short, the pellet was resuspended in 20 ml of extraction buffer (Tris-HCl 25 mM, pH 8; EDTA 0,1 mM; 1,5% Sarkosyl) and incubated 30 to 45 min at 4°C with constant stirring. The

extract was then centrifuged at 12000 g for 10 min at 4°C, the supernatant was equilibrated in 1% Triton and applied to a 5 ml glutathione-Sepharose 4B (Pharmacia) column. The fusion protein was purified according to the supplier's recommendations. Fractions of 2 ml were collected and proteins were quantified with a Bradford colorimetric assay (Bio-Rad). Antibodies were obtained by inoculating the recombinant protein (100 to 200 μg for each inoculation) to a rabbit using standard methods.

Whole fly extracts were prepared as follows : flies were homogenized in buffer A (10 mM HEPES, pH 7,9; 1 mM $CaCl_2$; 0,3 M sucrose; 0,5 mM DTT; 10 mM KCl; 0,5 mM PMSF) by 20 strokes in a loose fitting Dounce homogenizer. The extract was centrifuged for 5 min at 3000g over a fine mesh to avoid recovering large debris. The pellet was discarded and proteins in the supernatant were quantified with a Bradford assay.

Hemolymph was extracted from flies using the Nanoject apparatus (DrummondTM Scientific) and these extracts were recovered in PBS. Protein concentrations were determined with a Bradford assay.

All extracts (typically 100 μg of whole fly extract and 5 μg of hemolymph extract) were equilibrated in 2x Laemmli solution and denatured at 95°C for 1 min prior to loading on a 10% SDS-polyacrylamide gel. Following SDS-polyacrylamide gel electrophoresis, proteins were blotted to Hybond ECL nitrocellulose membranes (Amersham). Blots were incubated overnight at 4°C with a 1/5000 dilution of the GST-Spn43Ac polyserum. After washing with TBS (Tris Buffered Saline), the blots were incubated for 1 hour at 37°C with a 1/5000 dilution of HRP (Horse Radish Peroxidase)-conjugated donkey anti-rabbit secondary antibody (Amersham). The blots were developed using the ECL (Enhanced ChemiLuminescence) system

and x-ray film to detect the signal.

Infection procedures:

Bacterial challenge was performed by pricking adults with sharpened needles previously dipped into a concentrated bacterial culture of *E. coli* and *Micrococcus luteus*.

Natural infection by the entomopathogenic fungus *Beauveria Bassiana* was performed as follows. Anaesthetized flies were shaken for a few minutes in a Petri dish containing a sporulating fungal culture. Flies covered by spores were then removed to fresh *Drosophila* medium and incubated at 29°C.

Rescue experiments :

Spn43Ac protein was expressed via the *UAS-Spn43Ac/GAL4* system that allows the selective activation of any cloned gene in tissue-specific patterns [Brand, 1993, Development 118, 401]. Using the *EcoRI* site, filled in by the DNA polymerase Klenow fragment, and the *XhoI* site of the pUAST transformation vector (Brand, 1995 Curr. Opin. Neurobiol. 5(5):572), an approximately 1600 bp fragment corresponding to the *Spn43Ac* coding sequence was inserted 3' to the GAL4 UAS control element. In a similar cloning experiment, the *EcoRI* site, filled in by the DNA polymerase Klenow fragment, and the *XbaI* site of the pUAST transformation vector were used to insert the *Spn43Aa* coding sequence 3' to the GAL4 UAS control element. Transformant flies were obtained by microinjection, as described in Rubin, 1982, Science 218, 348, using a *w* recipient strain. Three and seven different transformant lines were obtained respectively for the *P(w^+ UAS-Spn43Ac$^+$)* and *P(w^+ UAS-Spn43Aa$^+$)* transgenes. Line 933 carrying the *P(w^+ UAS-Spn43Ac$^+$)* transgene on chromosome III and line 932 carrying the *P(w^+ UAS-Spn43Aa$^+$)* transgene on chromosome III were used in all experiments. The *nec^1/nec^2* allele

combination was chosen for rescue. A GAL4 line, in which the expression of the bacterial transactivator GAL4 is under the control of the ubiquitous promoter of the *daughterless* gene, was used for driving either the *P(w⁺ UAS-Spn43Ac⁺)* or the *P(w⁺ UAS-Spn43Aa⁺)* transgene. In this GAL4 line, the *P(w⁺ pda-GAL4)* transgene is inserted on the second chromosome and was recombined with the nec^2 allele in order to obtain nec^2, *P(w⁺ pda-GAL4)/CyO* flies. In parallel, both serpin transgenes were crossed into the *nec1/CyO* background and nec^1/*CyO; P(w⁺ UAS-Spn43Ac⁺)* or *P(w⁺ UAS-Spn43Aa⁺)/+* flies were selected. The *UAS* and *GAL4* stocks were crossed with each other to yield a heteroallelic nec^- combination. The resulting rescued nec^- animals, nec^2, *P(w⁺ pda-GAL4)/*nec^1*; P(w⁺ UAS-Spn43Ac⁺)* or *P(w⁺ UAS-Spn43Aa⁺)/+*, are Cy^+ and dark red-eyed and were distinguishable from nec^- siblings with only the *P(w⁺ pda-GAL4)* transgene.

Example 1

As described in detail herein, the present inventors have investigated the finding that the *drosomycin* gene is constitutively expressed in *nec* loss-of-function mutants. As discussed in the background to the invention, above, one of the hallmarks of the immune response of *Drosophila* is the induction of several genes encoding mostly small-sized and cationic peptides with potent anti-microbial activities. These peptides, which are predominantly but not exclusively produced in the fat body, fall into two broad categories according to their microbial targets: (I) anti-fungal peptides, namely drosomycin, and to a lesser extent metchnikowin, which is also active against Gram-positive bacteria, and (ii) anti-bacterial peptides, which include the cecropins, diptericin, attacin, drosocin and insect defensin.

To examine the expression of these immune markers in *nec*

loss-of-function mutants, Northern blots were prepared from total RNA of control and bacteria-challenged adults and probed with cDNAs corresponding to drosomycin, diptericin, cecropin A1 and metchnikowin. As could be expected from the inventor's previous studies, all genes were induced 6 hours after challenge in wild-type flies (Figure 4A). Remarkably, in *nec* loss-of-function mutants, the *drosomycin* gene was strongly expressed in the absence of immune challenge. Constitutive expression was also observed in the case of the *metchnikowin* gene, although it was less intense than that of *drosomycin*. In contrast, none of the anti-bacterial peptide genes was constitutively expressed in *nec* mutants. They were, however, induced by immune challenge as strongly as in wild-type flies, whereas the expression of *drosomycin* and *metchnikowin* was clearly enhanced by challenge over the level of constitutive expression observed in this mutant background.

Example 2

Expression of a functional *serpin Spn43Ac* gene in *nec* loss-of-function mutants rescues the phenotype of constitutive *drosomycin* gene expression

Previous experiments (Green *et al.*, in preparation) had shown that the *necrotic* phenotype was rescued by the expression of the *Spn43Ac* gene. To confirm that the constitutive expression of the *drosomycin* gene in *nec* mutants is due to the mutations in the same gene, the rescue experiments were extended, and the expression of the *drosomycin* gene was analysed in P-element insertion transgenic fly lines carrying the *Spn43Ac* coding sequence under the control of the yeast UAS sequence. These flies were crossed with flies expressing the yeast GAL4 activator under the control of the ubiquitously expressed *Drosophila daughterless* (*da*) gene promoter. To distinguish between

potential specific rescue of the *nec* phenotype by the *Spn43Ac* gene product and non-specific rescue by over-expression of an irrelevant *serpin*-encoding gene, transgenic flies were generated carrying another *serpin* gene of the *nec* locus, namely the *Spn43Aa* gene (see below). The expression of *drosomycin* was examined by Northern blotting analysis in both types of transgenic fly lines in the *nec* background and the results are shown in Figure 4B. The over-expression of the *Spn43Ac* gene in transgenic flies completely abolished the constitutive expression of *drosomycin*. These experiments also confirmed that *nec* fly lines carrying a functional *Spn43Ac* transgene do not exhibit necrotic spots on their bodies, as observed by Green and associates (in preparation). Importantly, though, the expression of the *Spn43Aa* gene was found not to rescue either the constitutive expression of the *drosomycin* gene, or the appearance of necrotic spots. See also Example 7.

Example 3

Spn43Ac functions upstream of the Toll regulatory gene cassette.

The above studies indicate that *Spn43Ac* is involved in the regulation of expression of the anti-fungal peptide genes, which is controlled by the *Toll* pathway [Lemaitre, 1996 Cell 86: 973-983]. To validate this assumption, the inventors analysed the expression of *drosomycin* in *nec; Tl*$^-$ and *nec; spz* double mutants. The constitutive expression of the *drosomycin* gene was abolished when *nec* alleles were combined with *Tl* or *spz* homozygous loss-of-function mutations (Figure 6A). Furthermore, when *gd* and *snk* loss-of-function alleles were combined with the *nec* mutation (Figure 6A) constitutive expression of the *drosomycin* gene was still evident, confirming that these proteases are not required for the induction of the anti-fungal response *via*

the *Toll* pathway.

As shown in earlier reports from one of the present groups, *drosomycin* expression is largely independent of the *imd* (for immune deficiency) pathway, which plays a crucial, although as yet undefined, role in the induction of the anti-bacterial peptides. The present inventors have also analysed the expression of *drosomycin* in *nec; imd* double mutants and found that the constitutive expression was strictly maintained in this background (Figure 6B).

Altogether these findings indicate that the function of *spz* and *Tl* is epistatic to *Spn43Ac*, and that *Spn43Ac* acts upstream of the *spz/Tl/pll* gene cassette. The data also confirm that the *Spn43Ac* gene does not function in the *imd* pathway and that the induction of the anti-bacterial peptide genes is independent of the proteolytic cascade(s), which are controlled by the *Spn43Ac* gene product.

Example 4

The *Spn43Ac* gene is upregulated by septic injury, fungal infection and in *Toll* gain-of-function mutants.

As stated above, *Toll* and all members of the downstream signalling cascade are expressed in adult *Drosophila* and their expression is upregulated following immune challenge [Lemaitre, 1996 Cell 86: 973-986]; for review, see Hoffmann, 1997 Trends Cell Biol. 7(7):309. This analysis was extended to the *Spn43Ac* gene. The data are presented in Figure 5A and, although some variability was observed between different experiments, they convincingly show that: (I) *Spn43Ac* is expressed in adults and is upregulated 2- to 6-fold by septic injury; (ii) in *Toll*-deficient mutants, expression of *Spn43Ac* is dramatically affected; the level of expression, even after septic injury, is at least two times lower than the constitutive expression in wild-type flies; (iii) in *Toll* gain-of-function mutants, the level of

the constitutive expression of *Spn43Ac* is 2 to 3 fold higher than the constitutive expression in wild-type flies; (iv) the expression of *Spn43Ac* is not affected in an *imd* mutant background.

Flies have further been subjected to an infection by the entomopathogenic fungus *Beauveria bassiana* by spraying the insects with fungal spores. Previous results had shown that this treatment, which mimics a natural infection, strongly induces expression of the anti-fungal peptide gene *drosomycin* and, to a lesser extent, that of *metchnikowin*, but fails to induce expression of the strictly anti-bacterial peptide genes [Lemaitre, (1997) P.N.A.S. USA 94, 14614]. Significantly, it has been observed here that this type of infection strongly induces expression of the *Spn43Ac* gene after a time lapse of 24 to 48 hours (Figure 5B).

It is concluded from these results that *Spn43Ac* is an immune-responsive gene and that its inducible expression, like that of *drosomycin*, is under control of the *Toll* signalling pathway.

Example 5

The *Spn43Ac* serpin functions in the hemolymph of Drosophila adults.

The genetic data presented above indicate that the serine protease inhibitor *Spn43Ac* functions upstream of the *Toll* signalling cascade. One obvious possibility in the present context is that *Spn43Ac* negatively regulates a humoral proteolytic cascade which eventually leads to the cleavage of Spaetzle to its active low molecular weight form. For that, the mature serpin must necessarily be present in the hemolymph of *Drosophila*. The predicted amino-acid sequence of Spn43Ac (see Figure 9) indeed contains a N-terminal stretch of hydrophobic amino-acids corresponding to a

putative signal peptide, which suggests that Spn43Ac is a secreted protein. A rabbit polyserum has been raised against a GST-Spn43Ac fusion protein and analysed by Western blot hemolymph samples from wild-type and mutant flies. Wild-type unchallenged flies were found to contain one immunoreactive doublet band at around 60 kDa; after immune challenge a second doublet band was observed at around 52 kDa (Figure 7). As both immunoreactive doublet bands were absent from hemolymph extracts of flies carrying a genomic deficiency that uncovers the *Spn43Ac* gene, it is deduced that they actually correspond to the *Spn43Ac* gene product. The size of the mature Spn43Ac protein calculated from the cDNA sequence is somewhat lower (52 kDa) than the sizes of the immunoreactive bands, which can be explained by post-translational modifications as serpins are generally glycoproteins [Potempa, (1994) J. Biol. Chem. 269; 15957-15960]. The appearance of the immune-induced 52 kDa band could correspond to the serpin cleaved by activated proteases, although this remains to be confirmed by appropriate analyses.

Example 6

Proteolytic cleavage of the Spaetzle protein is induced by immune challenge and is constitutive in *nec* mutants

To directly visualise the potential cleavage of the Spaetzle protein during the immune response analysis by Western blotting experiments has been made of protein extracts from wild-type and bacteria-challenged flies. A polyclonal antiserum directed against recombinant C-terminal Spaetzle has been generated and described by DeLotto and DeLotto [DeLotto, (1998) Mech. Dev. 72:141]. The antiserum recognises specifically recombinant Spaetzle produced in a baculovirus expression system. Importantly, the authors showed that the serine protease Easter cleaves

the recombinant protein at a unique position yielding a smaller carboxyterminal fragment of 12 kDa which is able to rescue the dorsalised phenotype when injected in the *spz* null females [DeLotto, (1998) Mech. Dev. 72:141]. In the present experiments, a clear signal corresponding to a protein of approximately 45 kDa in size was detected in denatured extracts of unchallenged flies. One hour after immune challenge, the 45 kDa band had disappeared whereas an immunoreactive doublet band of approximately 16 to 18 kDa had become apparent (Figure 8). These results suggest that the smaller protein corresponds to the processed form of the Spaetzle protein. The discrepancy between the sizes of the bands observed here and that reported by De Lotto and De Lotto can be explained by the fact that the present inventors were operating with total extracts of adult flies whereas De Lotto and De Lotto used recombinant protein. There are indications that the Spaetzle protein is glycosylated *in vivo* [Morisato, (1994) Cell 76, 677]; [DeLotto, (1998) Mech. Dev. 72:141].

Importantly, in the present context, the 16-18 kDa fragments were detected in extracts of unchallenged *nec* flies, together with the 45 kDa protein presumed to correspond to uncleaved Spaetzle. This result is in good agreement with the inventor's working hypothesis that the 16-18 kDa immunoreactive protein fragments correspond to the cleavage products of Spaetzle and that this cleavage is constitutive in *nec* mutants. Finally, the presence of significant amounts of the 45 kDa form of Spaetzle together with the 16-18 kDa fragments confirms at the protein level that, as for many other members of the *Toll* signalling pathway, the transcription of the *spaetzle* gene is regulated by a positive feedback loop.

Example 7

Materials and Methods

Drosophila stocks: The *nec* alleles, *Df(2R)sple-D1* and *Df(2R)sple-D2* chromosomes used in this example were from Heitzler *et al.* (1983). *Df(2R)pk-78k* and *Df(2R)nap-2* are from Gubb and Garcia-Bellido (1982) and Ringo *et al.* (1991), respectively. The region containing the *nec* gene is defined within the overlapping deletions *Df(2R)sple-D1* (43A1.2;43B2) and *Df(2R)nap-2* (41F4-9;43A1.2).

Crosses: Putative G1 transformant progeny were crossed to *w; In(2LR)O, Cy* dp^{lvI} *pr* cn^2*/Sco; In(3LR)TM2,* emc^2 Ubx^{130} e^s*/In(3LR)TM6B, Hu e Tb ca (w; CyO/Sco; TM2/TM6), or w; In(2LR)O, Cy* dp^{lvI} *pr* cn^2 *P{*ry^+ *wg:lacZ}/If; In(3LR)TM6B, Hu e Tb ca/Tp(3;3)MKRS, M(3)76A kar* ry^2 *Sb (w; CyO/If; MKRS/TM6B)* flies to establish balanced stocks in a w background. To test for rescue, third chromosomal *P{Spn43A}* inserts for each of the three serpin transcripts were crossed to nec^1 *bwD/CyO* flies. In the next generation, *w;* nec^1 *bwD/Sco; P{Spn43A}/+* or *w;* nec^1 bw^D*/If;P{Spn43A}/+)* males were crossed to *y w; Df(2R)pk-78k/CyO* females and the phenotype of surviving *w;* nec^1 bw^D*/Df(2R)pk-78k P{Spn43A}/+* progeny was scored. Flies were cultured at 25°C on yeasted cornmeal agar medium.

Cloning and sequencing:

Standard molecular biological techniques were used (Sambrook et al., 1989). Genomic inserts were isolated from the EMBL3 library of John Tamkun. cDNA clones were isolated from the plasmid 3rd instar imaginal disc library of Brown and Kafatos (1988) and lnm 1149 larval and adult head phage-insert libraries of Russell and Kaiser (1993). Spn43Aa cDNAs were isolated from the imaginal disc library using the 3.25 kb EcoR1 fragment from phage FP11/3; while Spn43Ab and Spn43Ac cDNAs were isolated from larval and head libraries using the using FP10/2 3.2 kb Sal1 and

FP10/2 2.1 + 6.0 kb Sal1 fragments, respectively (Figure 3). The sizes of the longest cDNA inserts were checked with transcript length estimates from Northern blots and the putative full length cDNAs (NB3, SL2 and SH8) were subcloned into pBluescript SK+ (Stratagene Ltd.) and restriction mapped. Smaller fragments of the cDNA inserts were subcloned and sequenced on both strands using Terminator Ready Reaction Mix (Perkin Elmer). Gaps in the sequence were filled in using specific oligonucleotide primers. The genomic sequence of the region was produced using similar methods. Comparison of the cDNA and genomic sequences identified the location and size of the introns.

Analysis of sequence data:

This was carried out using the Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, Wisconsin. DNA sequences for each transcript were compared to database sequences using the Blast program (Altschul *et al.*, (1990) J. Mol. Biol. 215: 403-410). The ClustalW program and SeqVu (Garvan Institute of Medical Research) were used for sequence alignment (Table 1) and identification of potential reactive centres.

Northern hybridisation:

Total and poly A+ RNA extractions and Northern blotting experiments were performed as described in Lemaitre *et al.*, (1996) Cell 86: 973-983. Probes corresponding to the cDNA of *Spn43Aa, Spn43Ab* and *Spn43Ac* were amplified by PCR using internal specific primers. 5mg of poly A+ RNA were loaded for each point. The probe for the ubiquitously expressed ribosomal protein *rp49* gene was used as a loading control (O'Connell, P. and Rosbach, M., (1984), Nucl. Acids Res. 12, 5495-5513).

Tissue *in situ* hybridisation: *In situ* staining followed the method of Tantz and Pfeifle (1989), *Chromosome* 98: 81-

85. The probes used were gel-purified inserts of the full length *Spn43A* cDNA clones; the 1.3 kb EcoR1 fragment of NB3 for *Spn43Aa*, the 1.3 kb *EcoR1* fragment of SL2 for *Spn43Ab* and the 0.6 kb + 0.7 kb *EcoR1* fragments of SH8 for *Spn43Ac*.

Transformation of flies:

Genomic constructs of each of the three serpins were made using the pWhiteRabbit transformation vector (Dunin-Borkowski and Brown, 1995). A 3.2kb *EcoRI* restriction fragment spanning the *Spn43Aa* restriction unit and a 7kb XhoI-BamHI fragment covering *Spn43Ac* were cut from the FP11/3 and FP10/2 genomic insert phage (Figure 3); the 5.2kb EcoRI fragment for *Spn43Ab* was cut from a cosmid (Cos 9/5) isolated from the cosmid library of J. Tamkun. A solution of 1mg/ml of one construct and 0.25mg/ml of the helper plasmid *ppi25.7wc* in Spradling buffer was microinjected into yellow white (*y w*) embryos following standard methods (Spradling, 1986,pp. 176-197 in Drosophila; A Practical Approach, edited by D. B. ROBERTS. IRL Press, Oxford).

RESULTS:

Three serpin transcripts have been identified within a 10 kb region in 43A1.2. Overlapping deletions that remove all three transcripts, *Df(2R)sple-D1/Df(2R)nap-2*, or the two most distal transcripts, *Df(2R)sple-D2/Df(2R)nap-2*, express amorphic *pk* and *nec* mutant phenotypes, but are otherwise wild type. The 5' exon of the *pk* gene maps between the most proximal serpin, *Spn43Aa*, and *Spn43Ab*. (D. Gubb, C. Green, D. Heun, D. Coulson, G. Johnson, S. Collier and J. RooteE, in preparation). See Figure 3.

Nucleotide and deduced amino acid sequences:

The cluster of three short transcripts in 43A1.2 shows homology to known serpins. The most proximal cDNA, *Spn43Aa*, is 1300 nucleotides long, Spn43Ab is 1333 nucleotides and

Spn43Ac is 1523 nucleotides. The longest open reading frame for each serpin is 370, 394 and 477 amino acids respectively. *Spn43Ac* has two introns while *Spn43Aa* and *Spn43Ab* each have three, as shown in Figure 3.

Temporal expression patterns:

Spn43Aa is expressed predominantly in the pupa up to two days, and also at a lower level in the embryo and in late larval stages. *Spn43Ab* is seen in the larval stages then again in late pupae and the adult. *Spn43Ac* is expressed at all stages from late embryo to adult.

Spatial expression patterns:

Spn43Ab gives concentric rings in the leg disc with a central dot at the position of the presumptive tarsal claw and is expressed after the morphogenetic furrow in the eye. Spn43Ac is expressed at the sites of innervated bristles on the notum and wing.

Rescue of phenotype:

Results of the test crosses are shown in Table 2. For the *P{Spn43Aa}* and *P{Spn43Ab}* crosses, phenotypically *nec* flies hatched, although at the reduced frequencies relative to Cy or Sco flies of 13% (89/646) and 6% (5/88), respectively. The surviving *nec*^{1}*/Df(2R)pk-78k* flies developed necrotic patches within 24 hours and died within three days of eclosion. In contrast, one eighth (96/745) of the *bwD w+ Cy+ If+* progeny of the *P{Spn43Ac}* test cross survived, together with 7.5% (56/745) of their *w nec Cy*$^{+}$ *If*$^{+}$ siblings. When reexamined ten days later, the *nec1 bw*^{D}*/pk-78k; P{Spn43Ac}/+* flies remained wild type, indicating that the *P{Spn43Ac}* insert rescues the necrotic phenotype completely.

Discussion

The present studies have focussed on flies which carry a mutation in a gene encoding a blood serine protease

inhibitor. As a result of this mutation, two striking phenotypes are observed. One is abnormal melanization at various sites throughout the body of the fly. This aspect, which led to the "necrotic" name given to the mutation, has not been further addressed in this study. The second phenotype, the constitutive expression of the gene encoding the antifungal peptide drosomycin, was the object of the current investigations.

Two mutations have been described so far which lead to constitutive expression of *drosomycin* (or, to a lesser extent, of *metchnikowin*). One is a gain-of-function mutation in the Toll transmembrane receptor, and the other is a loss-of-function mutations in the IκB-homologue Cactus. Interestingly, as is shown herein, the *nec* mutation affects a gene, *Spn43Ac*, which is not part of the Toll-Cactus intracellular pathway, but functions upstream of the Toll receptor.

The present data show that the constitutive expression of *drosomycin* due to a mutation in a blood serpin is mediated via Spaetzle and Toll and involves the appearance in the blood of a cleaved form of Spaetzle, which is likely to act as a ligand of Toll. The most compelling explanation is that the blood of *Drosophila* contains a protease which specifically cleaves the Spaetzle protein to its low-molecular active form. It may be assumed that this protease, which is probably at the downstream end of an amplification cascade of protease zymogens, exhibits a constant, although low, level of autocatalytic activation, which is a common feature of protease zymogens. It is also assumed that, under normal conditions, the autoactivated molecules are kept in check by the serine protease inhibitor Spn43Ac. Activation of Spaetzle and subsequent induction of the Toll pathway and expression of drosomycin

can occur in two situations: (I) immune challenge, which triggers the proteolytic amplification cascades and temporarily outbalances the capabilities of the serpin, thus inhibiting the activation of Spaetzle; (ii) chronic synthesis of a functionally deficient serpin which is unable to keep in check autoactivated Spaetzle-activating protease. The interpretation for (I) is substantiated by the observation that, after immune challenge, Spaetzle is massively cleaved to low molecular forms, and that this cleavage occurs constitutively in *nec* mutants. The validity of (ii) is convincingly illustrated by the abolition of the constitutive expression of *drosomycin* when a functional serpin gene is introduced under an ubiquitous promoter into a *nec* loss-of-function mutant background.

Conceptually, the activation of Spaetzle by a cascade of blood protease zymogens is remarkably similar to the coagulation response in the horseshoe crab *Tachypleus*. This response has been elegantly investigated at the biochemical and molecular levels by Iwanaga and associates (reviewed in Iwanaga, J. Biochem. (Tokyo) 123, 1 (1998)). It involves an upstream 123 kDa multidomain protein (Factor C), which associates complement control protein domains, EGF-like domains and a C-type lectin to a serine protease. Upon binding to bacterial lipopolysaccharide (LPS), the protease function is activated and a zymogen cascade, involving two additional serine proteases (Factor F and proclotting enzyme), is initiated. These two serine proteases contain compact domains with several disulphide bridges (CLIP domains) which are also present in the Snake and Easter serine protease precursors, which activate Spaetzle in the *Drosophila* embryo during dorso-ventral patterning. Significantly, the final target of the coagulation cascade in the horseshoe crab, i.e. the coagulogen protein, shows

tantalising structural similarities with the Spaetzle protein [Bergner, Biol. Chem. 378, 283 (1997)]. Three distinct serpins, which fall into the same general class as Spn43Ac, can specifically inhibit the proteases of the coagulation cascade.

Highly relevant for the present study is another proteolytic cascade- the prophenoloxidase cascade- which leads to melanization in arthropods. Here, a downstream oxidase, the phenoloxidase, is activated through a prophenoloxidase activating system which is a complex consisting of several proteases, protease inhibitors and recognition molecules that can recognise and respond to minute amounts of LPS from bacteria or β-1,3-glucan from fungi.

The results presented in the present specification, together with the parallels which have now been drawn with the coagulation and phenoloxidase cascades of non-*Drosophila* arthropods, lead to the concept that, in these cascades, non-self recognition is an upstream event. Toll does not qualify as a *bona fide* pattern recognition receptor in *Drosophila*, in contrast to what has been proposed for Toll-like receptor 2 in mammals. The actual pattern recognition receptor which initiates the cascade leading to the cleavage of Spaetzle and the activation of Toll remains to be identified. The *nec* mutants provided herein thus represent a welcome model for genetic screens to identify members of the activation cascade.

Of particular interest in the context of the immune response of *Drosophila* is furthermore the observation that serpin *Spn43Ac* gene is itself under the control of the Toll pathway. This indicates that activation of Toll by immune challenge will implicitly lead to increased production of an inhibitor of the Spaetzle activating cascade and thus

participate in shutting down the *spz/Tl/cact*-dependent transcription of *drosomycin*. Actually, this and most other immune-responsive genes in *Drosophila* are actively induced during the first hours following a challenge and their transcription slows down after 24 hours and in most cases has been shut off after 2 days.

Further, it is stressed that serpins are also involved in mammalian defence reactions, namely in blood coagulation, complement activation, fibrinolysis and inflammatory responses (Potempa, J. Biol. Chem., 1994, 269, 15957). Serpins are synthesised by the liver and are secreted into the blood where they represent the major class of serine protease inhibitors. They are also considered as strong acute phase reactants. Their role is critical in maintaining homeostasis and any mechanism reducing the functional level of members of this superfamily results in pathologies. In particular, genetic aberrations and deficiencies of serpin genes have been correlated with clinical syndromes such as pulmonary emphysema, angioedema and coagulopathies (Carrell et al., Nature, 1982, 298(5872):329). The present finding that a serpin is involved in the regulation of the *Drosophila* immune response therefore highlights once more the similarities between innate immunity in insects and mammals and reinforces the idea of a common ancestry.

The most closely related serpin to the *Spn43A* transcripts is from the tobacco hornworm, *Manduca sexta* (Jiang *et al.*, (1996) J. Biol. Chem. 271, 28017-28023) with 25 to 30% conservation at the amino acid level (Kanost *et al.*, 1989, J. Biol. Chem. 264, 965-972; Jiang et al., 1994, J. Biol. Chem. 269, 55-58; Jiang et al., 1996, J. Biol. Chem. 271, 28017-28023). Table 2 shows the level of homology between the deduced polypeptide sequences of various insect

serpins. This level of homology is characteristic within the serpin family, with the COOH-terminal half of the protein tending to show the greatest conservation (Sommer *et al.*, 1987, Biochemistry 26, 6407-6410). The protease-binding specificity of a particular serpin depends on the amino acid sequence of its reactive centre (Boswell and Carrell, 1988, Bioessays 8: 83-87; Huber and Carrell, 1989, Biochemistry 28: 8951-8966; Carrell and Evans, 1992, Curr. Opin. Struct. Biol. 2: 438-446).

Serpins control a number of proteolytic cascades in mammals and it may be that the black patches in *nec* mutants result from activation of the phenoloxidase cascade implicated in the wound healing response. The developmental pattern of expression of *Spn43Ab* at the segmental boundaries in the imaginal leg and behind the morphogentic furrow in the eye is particularly interesting with respect to possible proteolytic processes during morpogenesis. A *Rel/NFκB* signalling cascade has also been described in morphogenesis of mammalian limb bud growth (Kanegae et al., Nature 1998, 392(6676):611), which might well be controlled by a serpin/serine protease proteolytic cascade.

The lack of phenotype of deletions of the *Spn43Aa* and *Spn43Ab* transcripts may indicate either that the serpins are redundant under normal conditions, or that activation of their target proteases produces a subtle phenotype that does not affect viability under laboratory conditions. The Spn43Ab protein is highly basic (predicted isoelectric point 10), which suggests that it may have a DNA-binding function similar to that of the MENT serpin [Grigoryev, Bednar and Woodcock (1999), J. Biol. Chem. 274: 5626-5636]. The identification of the target proteases and the manipulation of the serpin/protease balance may well allow the biochemical pathways to be identified.

Acknowledgements

The present inventors are grateful to Dr. Robert DeLotto for generously providing anti-Spaetzle antiserum, Dr. J. Roote for the fly stocks, Dr. Julien Royet for critical reading of the manuscript. The excellent technical assistance of Anny Menier, Reine Klock and Marie-Eve Moritz is gratefully acknowledged. The inventors' work was supported by grants from the Human Frontiers in Science Program, the Training and Mobility of Researchers Program, and MRC Programme grants to M. Ashburner, D. Gubb and S. Russell.

References:

The references cited in the above description are all incorporated by reference in their entirety.

TABLE 1

Percent identity (similarity) of invertebrate serpins

	Spn43Ab	*Spn43Ac*	*Acp76A*	*Mand*
Spn43Aa	35 (46)	36 (46)	22 (30)	37 (46)
Spn43Ab		30 (38)	22 (34)	27 (40)
Spn43Ac			?	4 (44)
Acp76A				21 (30)

TABLE 2

Test crosses for rescue of nec phenotype

♂	x ♀	w Cy^+ Sco^+	bw^D Cy^+ Sco^+	(w) Cy (or, Sco)	(w) Cy Sco
w; nec^1 bw^D/Sco; P{Spn43Aa}/+	x *w; Df(2R)pk-78k/CyO*	55*	34*	383	174
		w Cy^+ If^+	bw^D Cy^+ If^+	(w) Cy (or, If)	(w) Cy If
w; nec^1 bw^D/If; P{Spn43Ab}/+	x *w; Df(2R)pk-78k/CyO*	3*	2*	59	24
w; nec^1 bw^D/If; P{Spn43Ac}/+	x *w; Df(2R)pk-78k/CyO*	56*	96**	417	176

* Developed necrotic patches and died within 72 hours of eclosion.

** Survived at least 10 days after eclosion with nec^+ phenotype.

CLAIMS

1. A substance which is an isolated polypeptide comprising an amino acid sequence as set out in Figure 9.

2. A substance which is an isolated polypeptide having greater than about 70% amino acid sequence identity with the amino acid sequence set out in Figure 9.

3. A substance which is an isolated polypeptide having greater than about 80% amino acid sequence identity with the amino acid sequence set out in Figure 9.

4. A substance which is an isolated polypeptide having greater than about 90% amino acid sequence identity with the amino acid sequence set out in Figure 9.

5. A substance which is an isolated polypeptide which is a mutant, variant, derivative or allele of the polypeptide of any one of claims 1 to 4.

6. A substance which comprises an active portion, fragment, chemical derivative or functional mimetic of the polypeptide of any one of claims 1 to 5, and which retains a biological property of a native serpin protein.

7. A substance according to claim 6 wherein said active portion, fragment, derivative or functional mimetic comprises a serine protease-binding site.

8. A substance according to claim 7 wherein said serine protease-binding site is from a serpin.

9. A substance which is a polypeptide or peptide of any one of the preceding claims, which has a modulated or altered biological activity compared with a native serpin protein.

10. A substance according to claim 9 which has an impaired ability to inhibit serine protease activity compared with a native serpin protein.

11. An isolated nucleic acid having a nucleotide sequence which encodes a polypeptide or peptide according to any one of the preceding claims.

12. An isolated nucleic acid according to claim 11 comprising a nucleotide sequence as set out in Figure 2.

13. An isolated nucleic acid according to claim 11 having greater than about 50% sequence homology with the sequence of Figure 2.

14. An isolated nucleic acid according to claim 11 having greater than about 60% homology with the sequence of Figure 2.

15. An isolated nucleic acid according to claim 11 having greater than about 70% homology with the sequence of Figure 2.

16. An isolated nucleic acid according to claim 11 having greater than about 80% homology with the sequence of Figure 2.

17. An isolated nucleic acid according to claim 11 having

greater than about 90% homology with the sequence of Figure 2.

18. An isolated nucleic acid according to any one of claims 11 to 17, having one or more mutations which result in the loss or reduction of a biological activity of the encoded polypeptide or peptide compared with native serpin protein.

19. An isolated nucleic acid according to claim 18 wherein a said mutation comprises a nucleotide deletion, substitution, or insertion.

20. An isolated nucleic acid according to claim 19 wherein a said mutation comprises a deletion as set out in Figure 3.

21. An antisense oligonucleotide which has a nucleotide sequence complementary to a nucleic acid of any one of claims 11 to 20.

22. A vector or construct comprising an isolated nucleic acid or oligonucleotide molecule of any one of claims 11 to 21.

23. A host cell comprising a vector or construct of claim 22.

24. A host cell according to claim 23 which is a *Drosophila* cell.

25. Use of a host cell of claim 23 or claim 24 to produce a polypeptide or peptide of any one of claims 1 to 10.

26. Method for producing a polypeptide or peptide of any one of claims 1 to 10, comprising culturing a host cell of claim 23 or claim 24 under conditions suitable for expression of the polypeptide or peptide.

27. Method according to claim 26 further comprising isolating the polypeptide or peptide from the cell culture.

28. Use of a polypeptide or peptide of any one of claims 1 to 10 in a method of screening for test molecules which are capable of affecting or modulating the interaction between a serpin and a serine protease.

29. A substance which is capable of modulating the interaction between a serpin and a serine protease, as obtainable by the use according to claim 28 of a polypeptide or peptide in a said screening method.

30. A substance according to claim 29 which is an agonist of a polypeptide or peptide of any one of claims 1 to 10.

31. A substance according to claim 29 which is an antagonist of a polypeptide or peptide of any one of claims 1 to 10.

32. An antibody which is capable of binding a polypeptide or peptide of any one of claims 1 to 10.

33. An antibody according to claim 32 which is a monoclonal antibody.

34. Use of an antisense oligonucleotide of claim 21, a substance of any one of claims 29 to 31, or an antibody of

claim 32 or claim 33, to modulate or alter the expression in a cell of a polypeptide or peptide of any one of claims 1 to 10, or to modulate or alter the interaction between a serpin and a serine protease, respectively.

35. Method of reducing the inhibition of a serine protease, comprising causing the expression in a cell of a substance as claimed in claim 10.

36. Method of modulating or altering the downstream expression of a nucleic acid in a cell, comprising the step of reducing the inhibitory effect of a serpin on serine protease activity.

37. Method according to claim 36, wherein said nucleic acid encodes an anti-fungal or anti-viral peptide.

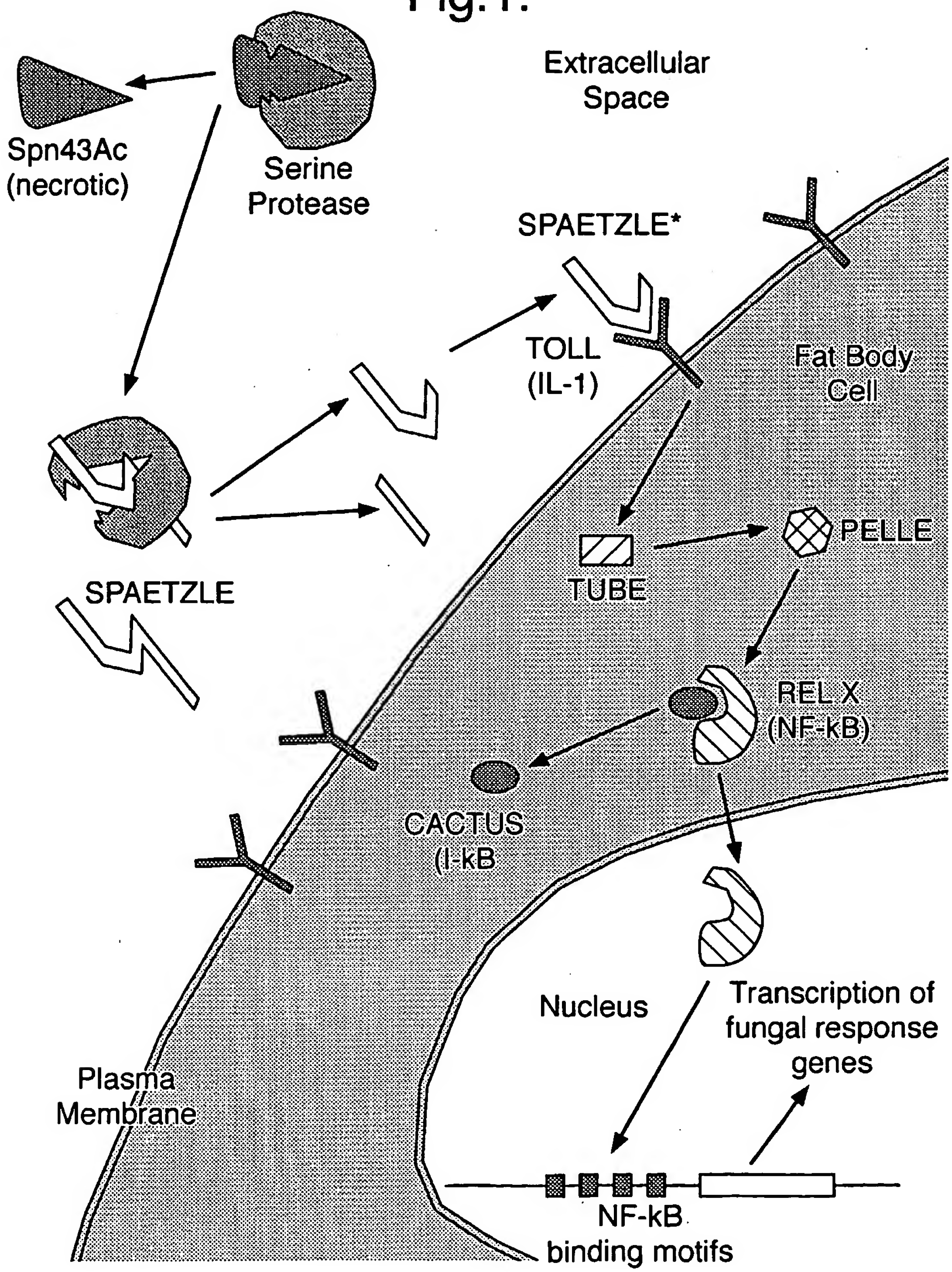
38. Method of manufacturing a peptide or polypeptide in a cell, comprising the step of causing an enhanced downstream expression of said peptide or polypeptide in the cell using a method according to claim 36.

39. Method according to claim 38, wherein said peptide is an anti-fungal or anti-viral peptide.

40. A substance which is a peptide or polypeptide molecule as obtainable using a method of claim 38 or claim 39.

41. Use of a peptide molecule according to claim 40 in the manufacture of a pharmaceutical composition.

Fig.1.


Extracellular
Space
Spn43Ac
(necrotic)
Serine
Protease
SPAETZLE*
TOLL
(IL-1)
Fat Body
Cell
SPAETZLE
TUBE
PELLE
REL X
(NF-kB)
CACTUS
(I-kB
Nucleus
Transcription of
fungal response
genes
Plasma
Membrane
NF-kB
binding motifs

SER1 (Spn43Aa) cDNA :

Fig.2.

AAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGACCT
GCACCCCCCCCCCCCGAGTCTCGAGGTGTCGACGCAGGCGCAATGAACCACTGG
CTAAGTATCATTCTTCTTGGGGTGTGGATATCCGCTCCTGAAGGTCTGGTAACA
CGATCAAGGATCGCAATCTCTTCGCCACCGAACTCTTCCAAACCCTCGCCACAGA
TCGCCAGGATGAGAACGTGATCATCTCGCCGGTTTCCATCCAGCTGGCCCTCGGG
TTGGCTTACTACGGAGCTGAGGGCAGGACGGCCGCGGAACTGCAGAAGACCTTG
CACGCCTCCGCCAAGGAGAGCAAAGATGGTCTGGCCGAGAGCTACCACAACCTG
CTGCACTCTTACATCAAGTCCAAGACCGTGTTGGAGATCGCCAACAAGGTGTACA
CCCGGCAGAATCTCACGGTATCTAGCCACTTCCGAGAGGTGGCCCAGAAGTACTT
CGACTCCGAGGTAGAACCACTGGACTTCAGTCGCGAAACGGAGGCCGTGGAGCA
GATCAACCGCTGGGTGAAGCAGCAGACGGAGAACAAGATCGAACGGGTGGTGG
AAAGCCTGGAGCCGGACACCAATGTGGCGCTGGTCAACGCCATCTACTTCAAGG
CTCGCTGGGCACGACCCTTCAACGACGAGGATACCCGGGATCGCGAGTTCTGGC
TGAGCGAGAGCCGGTCCATCCAGGTGCCAACCATGTTCGCAGACAACTGGTACT
ACTACGCCGACTACCCGGAACTCGACGCCAAGGCCATTGAGCTGTTCTTCGAAA
ACATCAACCTGACCATGTGGTTCATCCTGCCCAACCAGCGTTCCGGACTACAGGC
TCTGGAGCAGAAGCTCAAGGGCGTCGACTTCAATCTGCTCGAAGACCGCTGGCA
GTGGCAGAGTGTGTCCGTTTACCTGCCCAAGTTCAAGTTCGAGTTCGACACGGAC
CTAAGACCCACATTGCATAAGATGGGAATCAGTGCCATGTTCTCAGATGCAGCCG
ACTTCAGCAACATTTTCCAGGACTCGCCCATCGGCACTCGGATCACAAAGGTGCA
GCACAAGACCTTCATCGATGTGAACGAGATCGGATGTGAGGCGGCTGGAGCTAG
TTATGCTGCCGGAGTTCCCATGTCCCTGCCCCTGGACCCCAAGACTTTCGTGGCC
GATCATCCATTTGCGTTCATCATTCGCGACAAGCACGCTGTCTATTTCACCGGAC
ACATTGTCAAGTTTTAATCGTCACACATTTCCCTCAGATTAAGCACTTTAAATTGT
AATCATTACATCAATAAATAAATGCGGAGAGCCTCCGCCAGAACTTAAAAAAAA
AAAAAACGAATGCTGCGGCCGCCGAATTC

SER2 (Spn43Ab) cDNA :

GAATTCGGCTCAAGTGCGAATCAAGCGGAGcAGGAGAcGCCCATCAACAATTTAT
GTGACGAcCATGGCTGTCATCATCAGcTGcCTATTACTTCTCCTCGCGACAGTGTC
GCAGTCCAAGACCGTGGGTTACGATGCGGCTGCCGATCGCAaTTTGgTGGCTGcCG
ATCTCTACAACGCCGTGGXCGCCGATCATCTXAACGAAAATGTGGTCATCTCGCC
AGCGACCATACAGAGTTCCATGGCGCTGGCCTTCGTGGGGGGCAAGGGTCAGAC
GGcATCGGAGTTGCAGCAGGGGCTGCGTTTAGGTCCTGGCGATGCGGATGCGGTG
AGCCAGCGCAGCGGTAGTTACCAACAGGCCCTGACCCGCGACAACAACTTCCGG
CTGGCCAACAACATCTACATCAACGAGAACCTTGAGTTCAAGGGCTCCTTCAGG
GACGTGGCCCAGCGCCAGTTCGACTCGAACATCGACAAGCTGGACTTTCACCCG
CCGTACAACAAGCGCACGGcGGATGGcATCAACCGGgcGGTGGcCACCAAGACCA
ACGGcAAGATCACTGACATCCTCCGCGCCGAACTGCTGAATGATCGCACCGAGG
GAGTGATCGTGAaCGGcGTTTCCTACTCAGcCGCCTGGCAAAATGCCTTCCGGCTG
GACAAGACCGAAAAGCGCTCCTTCCGCACCGGAAGCGGGCAGTCCGTCAAGGTA
GACACCATGTGGACgctGCAGAACTTCAACTACGCCGAGGTCAACTCCTTGGACGC
CAAGGTGGTGGAGCTGCCCTACCAGAACCCCGACTTCTCCATGCTGCTGCTCCTG
CCCAATCGCAAGGATGGTCTGAGATCCCTGCAGCAGTCGCTGTCCGGTAAGAAT
CTT
CTGGCCGAAATCGGAGCGTTGAGCCAgCAAAAGGTGGAGGTGCTgCTGcCCAAGT
TTAGTGTCACCTTTGGCTTGGGTTTGGaGGGACCATTCAAGAAGCTGGGCGTCCA

Fig.2 (Cont).

TACAATGTTCTCGAGAGATGGTGACTTTGGcAACATGTACCGCATGTTTGTCAGC
CATTTCATTAACGCAGTGGAGCACAAGGCCAATGTGGAGGTCACTGAAGCTGGC
GTGGATCAACCCCTGGAAACTGGATTGCTTAAAGGACTCTTTTCGCGCTCCAAGA
AGTTCGAGGCAGATCATCCGTTCGTGTTCGCCATCAAGTACAAGGACTCAATCGC
CTTTATCGGACACATTGCTAACTATGCTTATGTCTAAGCTAGATCGCAGATACGG
AGTAGATCATTTTGTTAAAACTTACAAATAAAAAACACACGTAAATTGAAAAAA
AAAAAAACCGAATTC

SER3 (Spn43Ac) cDNA :

AATTCAAGCCATTCATTTAGAAGGTTTCCAGCTAGAAATAACAGATAAGAAACA
TGGCGAGCAAAGTCTCGATCCTTCTCCTGCTAACCGTCCATCTTCTGGCTGCTCA
GACCTTCGCCCAGGAGCTCATCGCTTGGCAACGGCAACAACAACAGCAGCAACA
GCAGCAACTGCAGCTACAGCAGCAACTGCTGCTGCAGCAGCAACAACACCAACG
TAACCCAAGACCGGAGCTGGGCCTCCGTTCCCTGCCCGGAAACCCGTGGACCCA
GAACAATCAGGAAGCCATAAGCGATGTGGTGGCGGTGGACCTAACCAAACGTGA
GCCGGTCACTCCGCCACCCAATCGCCCGCCGCCCGTCTTCAGCTACATGGACCGC
TTCAGCTCCGAGCTCTTCAAGGAGATCATTAAGTCGCAAAGTCAGCAGAACGTG
GTGTTCTCGCCCTTCTCCGTCCACGCGCTGCTGGCCCTGATCTACGGGGCCTCGG
ACGGAAAAACGTTCCGGGAACTGCAGAAGGCCGGAGAGTTCAGCAAGAACGCC
ATGGCCGTGGCCCAGGACTTCGAGAGCGTGATCAAGTACAAGAAGCATTTGGAG
GGCGCAGATCTGACCCTGGCAACCAAGGTCTACTACAACCGCGAGCTGGGTGGC
GTCAACCACAGCTACGATGAGTACGCAAAGTTCTATTTCAGCGCCGGCACGGAG
GCTGTCGACATGCAGAACGCCAAAGACACGGCAGCCAAGATCAACGCCTGGGTG
ATGGACACGACGCGGAACAAGATCCGGGACCTGGTCACACCGACCGACGTTGAC
CCACAGACGCAGGCGCTCCTTGTGAATGCAGTCTACTTCCAGGGTCGTTGGGAGC
ACGAATTCGCCACCATGGACACATCACCCTACGACTTTCAGCACACCAACGGCA
GAATTTCCAAGGTGGCCATGATGTTCAACGACGATGTGTACGGCCTGGCCGAGCT
GCCTGAGCTGGGCGCCACCGCCTTGGAACTGGCCTACAAGGACAGCGCCACCAG
CATGCTAATCCTGCTGCCCAATGAGACCACCGGACTGGGCAAAATGCTGCAGCA
GCTGTCCCGGCCGGAGTTCGATCTCAACCGCGTTGCCCACCGCCTGCGCCGCCAG
TCGGTCGCTGTGCGCCTGCCCAAGTTCCAATTCGAGTTCGAGCAGGACATGACCG
AGCCGCTAAAGAACCTGGGAGTCCACCAGATGTTCACGCCCAACTCGCAGGTGA
CCAAGTTGATGGATCAGCCGGTGCGCGTGAGCAAGATCCTGCAGAAGGCCTACA
TCAATGTGGGCGAGGCGGGCACAGAGGCCTCGGCAGCTTCCTATGCCAAGTTCG
TACCCCTTTCGCTGCCTCCCAAGCCCACGGAGTTCGTCGCCAACCGGCCATTCGT
CTTCGCCGTCCGCACCCCCTCCTCAGTTCTGTTCATAGGTCACGTGGAGTATCCCA
CGCCCATGAGCGTCTAAAATGAACCAAATTCGGACAAATTGATTTATTAGAATTC

4/9

Fig.3.

Map of the *Spn43A* region

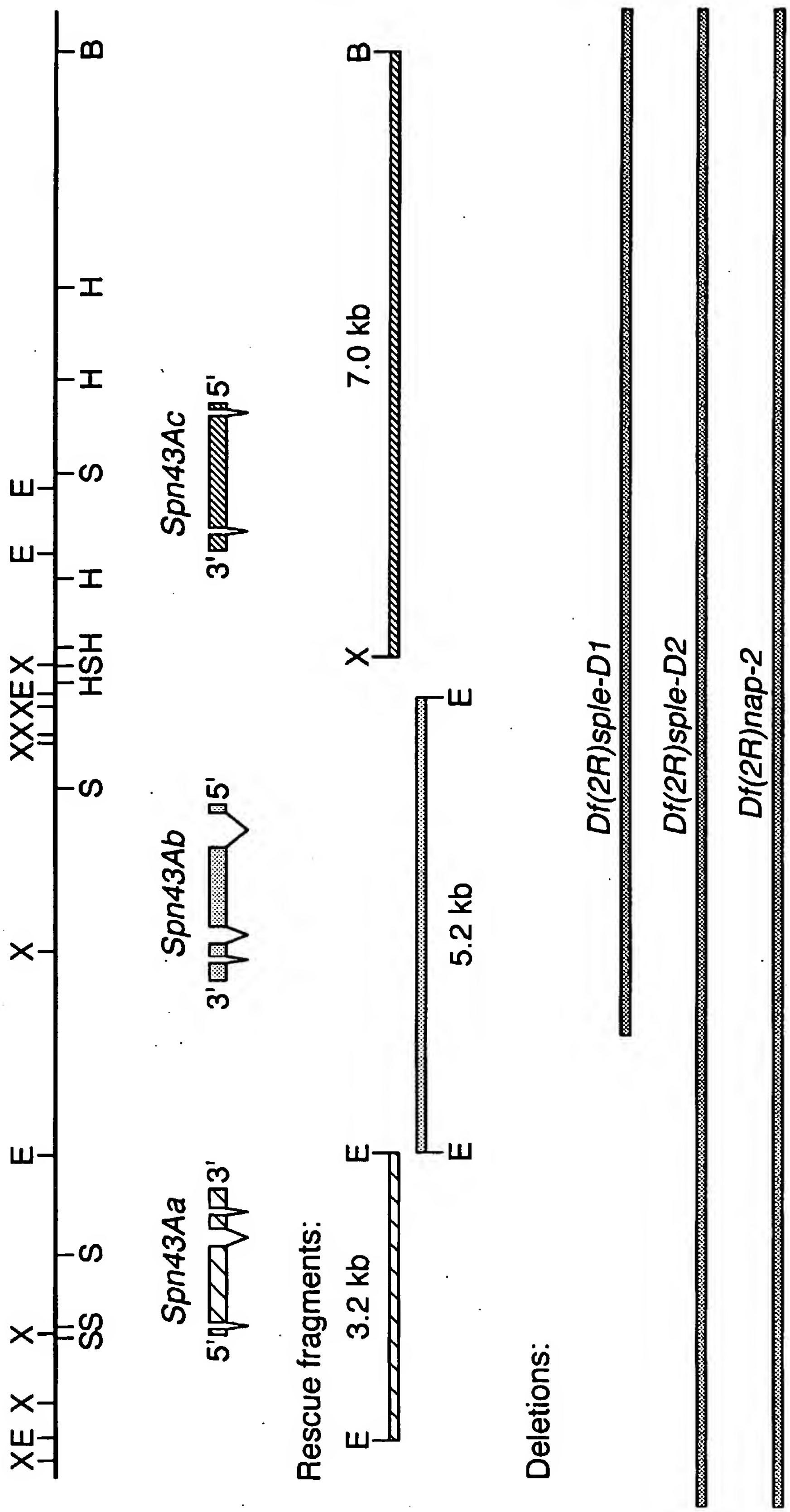


Fig.4A.

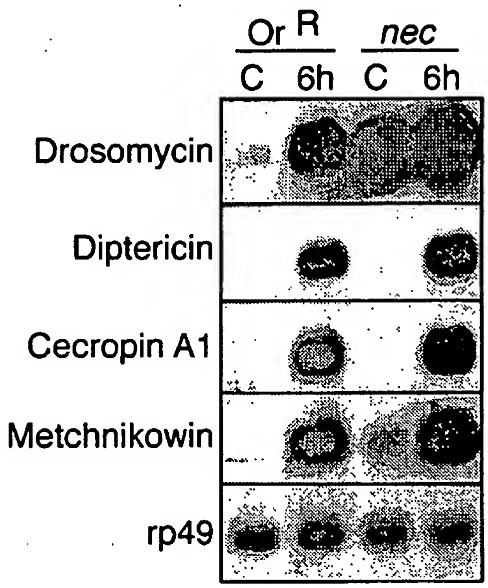

Or R
nec
C
6h
C
6h
Drosomycin
Diptericin
Cecropin A1
Metchnikowin
rp49

Fig.4B.

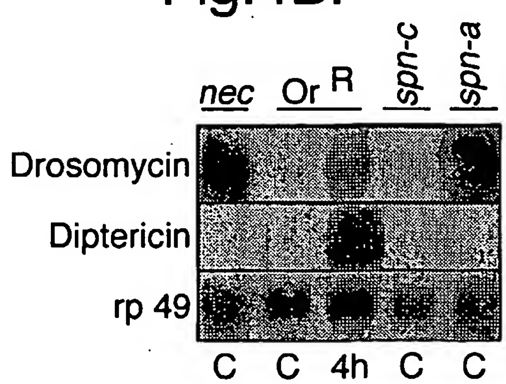

nec
Or R
spn-c
spn-a
Drosomycin
Diptericin
rp 49
C
C
4h
C
C

Fig.5A.

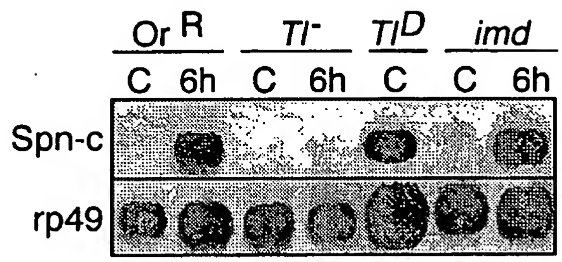


Fig.5B.

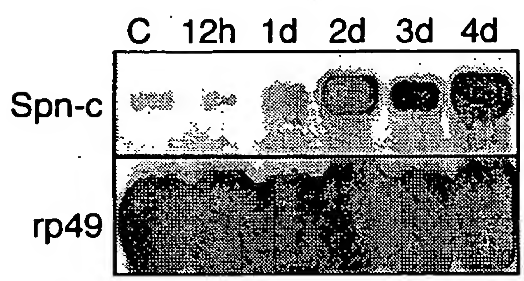


Fig.6A.

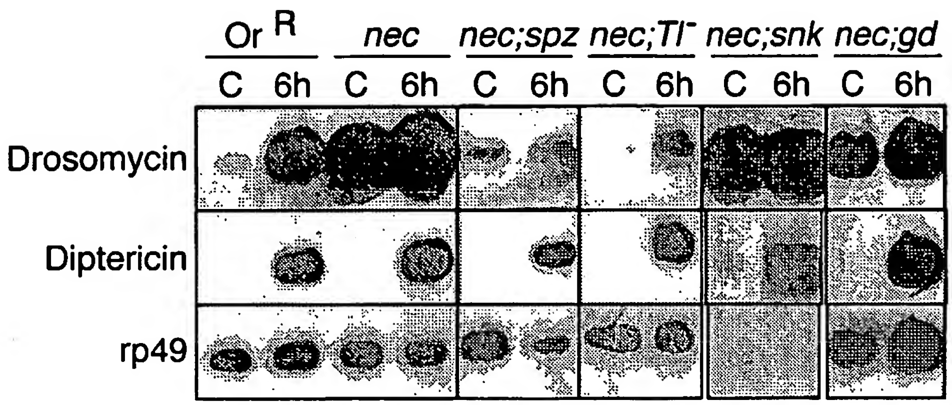

Or R
nec
nec;spz
nec;Tl⁻
nec;snk
nec;gd
C 6h C 6h C 6h C 6h C 6h C 6h
Drosomycin
Diptericin
rp49

Fig.6B.

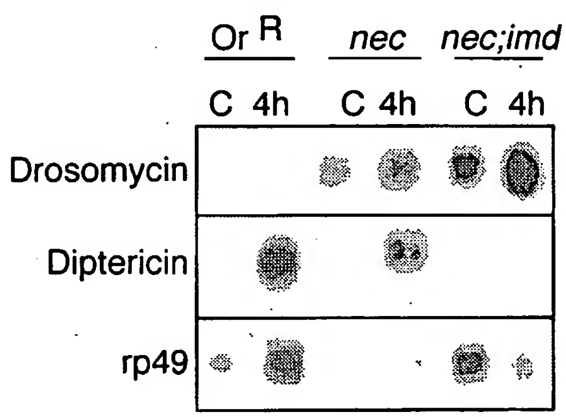

Or R
nec
nec;imd
C 4h C 4h C 4h
Drosomycin
Diptericin
rp49

Fig.7.

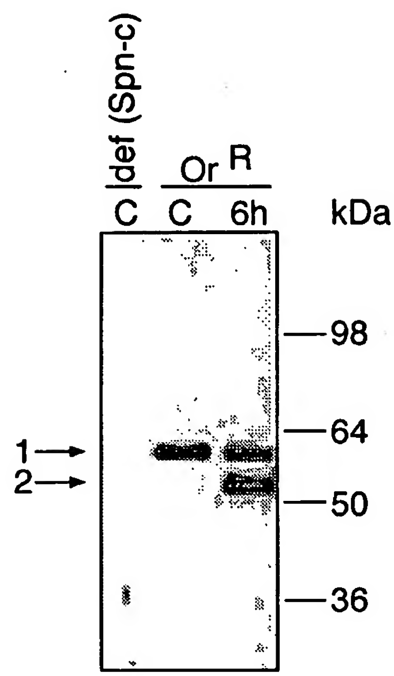


Fig.8.

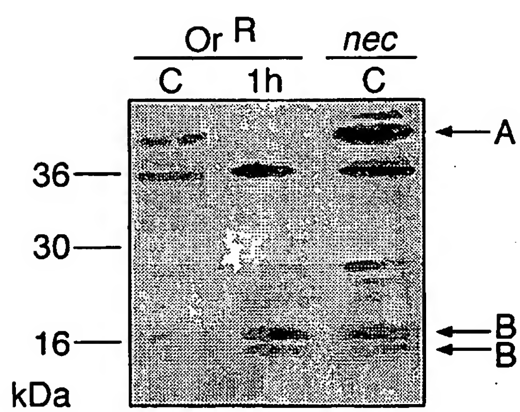


Fig.9.

SER1 (Spn43Aa) Protein :

MNHWLSIILLGVWISAPEGLGNTIKDRNLFATELFQTLATDRQDENVIISPVSIQLALG
LAYYGAEGRTAAELQKTLHASAKESKDGLAESYHNLLHSYIKSKTVLEIANKVYTR
QNLTVSSHFREVAQKYFDSEVEPLDFSRETEAVEQINRWVKQQTENKIERVVESLEP
DTNVALVNAIYFKARWARPFNDEDTRDREFWLSESRSIQVPTMFADNWYYYADYPE
LDAKAIELFFENINLTMWFILPNQRSGLQALEQKLKGVDFNLLEDRWQWQSVSVYLP
KFKFEFDTDLRPTLHKMGISAMFSDAADFSNIFQDSPIGTRITKVQHKTFIDVNEIGCE
AAGASYAAGVPMSLPLDPKTFVADHPFAFIIRDKHAVYFTGHIVKF

SER2 (Spn43Ab) Protein :

MAVIISCLLLLLATVSQSKTVGYDAAADRNLVAADLYNAVXADHLNENVVISPATIQ
SSMALAFVGGKGQTASELQQGLRLGPGDADAVSQRSGSYQQALTRDNNFRLANNIY
INENLEFKGSFRDVAQRQFDSNIDKLDFHPPYNKRTADGINRAVATKTNGKITDILRA
ELLNDRTEGVIVNGVSYSAAWQNAFRLDKTEKRSFRTGSGQSVKVDTMWTLQNFN
YAEVNSLDAKVVELPYQNPDFSMLLLLPNRKDGLRSLQQSLSGKNLLAEIGALSQQK
VEVLLPKFSVTFGLGLEGPFKKLGVHTMFSRDGDFGNMYRMFVSHFINAVEHKANV
EVTEAGVDQPLETGLLKGLFSRSKKFEADHPFVFAIKYKDSIAFIGHIANYAYV

SER3 (Spn43Ac) Protein :

MASKVSILLLLTVHLLAAQTFAQELIAWQRQQQQQQQQLQLQQQLLLQQQQHQR
NPRPELGLRSLPGNPWTQNNQEAISDVVAVDLTKREPVTPPPNRPPPVFSYMDRFSSE
LFKEIIKSQSQQNVVFSPFSVHALLALIYGASDGKTFRELQKAGEFSKNAMAVAQDFE
SVIKYKKHLEGADLTLATKVYYNRELGGVNHSYDEYAKFYFSAGTEAVDMQNAKD
TAAKINAWVMDTTRNKIRDLVTPTDVDPQTQALLVNAVYFQGRWEHEFATMDTSP
YDFQHTNGRISKVAMMFNDDVYGLAELPELGATALELAYKDSATSMLILLPNETTG
LGKMLQQLSRPEFDLNRVAHRLRRQSVAVRLPKFQFEFEQDMTEPLKNLGVHQMFT
PNSQVTKLMDQPVRVSKILQKAYINVGEAGTEASAASYAKFVPLSLPPKPTEFVANR
PFVFAVRTPSSVLFIGHVEYPTPMSV